(3)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 April 2003 (10.04.2003)

PCT

(10) International Publication Number
WO 03/029459 A2

(51) International Patent Classification[7]: C12N 15/11, A61K 48/00

(21) International Application Number: PCT/EP02/10881

(22) International Filing Date: 27 September 2002 (27.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

01123453.1	28 September 2001 (28.09.2001)	EP
02006712.0	22 March 2002 (22.03.2002)	EP
02016772.2	26 July 2002 (26.07.2002)	EP

(71) Applicant *(for all designated States except US)*: **MAX-PLANCK-GESELLSCHAFT ZUR FÖRDERUNG DER WISSENSCHAFTEN E.V.** [DE/DE]; Hofgartenstrasse 8, 80539 München (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TUSCHL, Thomas** [DE/DE]; Keplerstrasse 9, 37085 Göttingen (DE). **LAGOS-QUINTANA, Mariana** [MX/DE]; Kurze Geismarstrasse 13, 37073 Göttingen (DE). **LENDECKEL, Winfried** [DE/DE]; Blinde Gasse 52, 37318 Hohengandern (DE). **MEYER, Jutta** [DE/DE]; Sternstrasse 16, 37083 Göttingen (DE). **RAUHUT, Reinhard** [DE/DE]; Merkelstrasse 16, 37085 Göttingen (DE).

(74) Agents: **WEICKMANN, Franz, Albert** et al.; Weickmann & Weickmann, Postfach 860 820, 81635 München (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

US Serial No 10/490,955 filed 9/15/2004
03/17/05

WO 03/029459 A2

(54) Title: MICRORNA MOLECULES

(57) **Abstract:** In Caenorhabditis elegans, lin-4 and let-7 encode 22- and 21 -nucleotide RNAs, respectively, that function as key regulators of developmental timing. Because the appearance of these short RNAs is regulated during development, they are also referred to as "small temporal RNAs" (stRNAs). We show that many more 21- and 22-nt expressed RNAs, termed microRNAs, (miRNAs), exist in invertebrates and vertebrates, and that some of these novel RNAs, similar to let-7 stRNA, are also highly conserved. This suggests that sequence-specific post-transcriptional regulatory mechanisms mediated by small RNAs are more general than previously appreciated.

MicroRNA molecules

Description

The present invention relates to novel small expressed (micro)RNA molecules associated with physiological regulatory mechanisms, particularly in developmental control.

In Caenorhabditis elegans, lin-4 and let-7 encode 22- and 21-nucleotide RNAs, respectively (1, 2), that function as key regulators of developmental timing (3-5). Because the appearance of these short RNAs is regulated during development, they are also referred to as "microRNAs" (miRNAs) or small temporal RNAs (stRNAs) (6). lin-4 and let-21 are the only known miRNAs to date.

Two distinct pathways exist in animals and plants in which 21- to 23-nucleotide RNAs function as post-transcriptional regulators of gene expression. Small interfering RNAs (siRNAs) act as mediators of sequence-specific mRNA degradation in RNA interference (RNAi) (7-11) whereas miRNAs regulate developmental timing by mediating sequence-specific repression of mRNA translation (3-5). siRNAs and miRNAs are excised from double-stranded RNA (dsRNA) precursors by Dicer (12, 13, 29), a multidomain RNase III protein, thus producing RNA species of similar size. However, siRNAs are believed to be double-stranded (8, 11, 12), while miRNAs are single-stranded (6).

We show that many more short, particularly 21- and 22-nt expressed RNAs, termed microRNAs (miRNAs), exist in invertebrates and vertebrates, and that some of these novel RNAs, similar to let-7 RNA (6), are also highly conserved. This suggests that sequence-specific post-transcriptional

regulatory mechanisms mediated by small RNAs are more general than previously appreciated.

The present invention relates to an isolated nucleic acid molecule comprising:

(a) a nucleotide sequence as shown in Table 1, Table 2, Table 3 or Table 4

(b) a nucleotide sequence which is the complement of (a),

(c) a nucleotide sequence which has an identity of at least 80%, preferably of at least 90% and more preferably of at least 99%, to a sequence of (a) or (b) and/or

(d) a nucleotide sequence which hybridizes under stringent conditions to a sequence of (a), (b) and/or (c).

In a preferred embodiment the invention relates to miRNA molecules and analogs thereof, to miRNA precursor molecules and to DNA molecules encoding miRNA or miRNA precursor molecules.

Preferably the identity of sequence (c) to a sequence of (a) or (b) is at least 90%, more preferably at least 95%. The determination of identity (percent) may be carried out as follows:

$$I = n : L$$

wherein I is the identity in percent, n is the number of identical nucleotides between a given sequence and a comparative sequence as shown in Table 1, Table 2, Table 3 or Table 4 and L is the length of the comparative sequence. It should be noted that the nucleotides A, C, G and U as depicted in Tables 1, 2, 3 and 4 may denote ribonucleotides,

deoxyribonucleotides and/or other nucleotide analogs, e.g. synthetic non-naturally occurring nucleotide analogs. Further nucleobases may be substituted by corresponding nucleobases capable of forming analogous H-bonds to a complementary nucleic acid sequence, e.g. U may be substituted by T.

Further, the invention encompasses nucleotide sequences which hybridize under stringent conditions with the nucleotide sequence as shown in Table 1, Table 2, Table 3 or Table 4, a complementary sequence thereof or a highly identical sequence. Stringent hybridization conditions comprise washing for 1 h in 1 x SSC and 0.1% SDS at 45°C, preferably at 48°C and more preferably at 50°C, particularly for 1 h in 0.2 x SSC and 0.1% SDS.

The isolated nucleic acid molecules of the invention preferably have a length of from 18 to 100 nucleotides, and more preferably from 18 to 80 nucleotides. It should be noted that mature miRNAs usually have a length of 19-24 nucleotides, particularly 21, 22 or 23 nucleotides. The miRNAs, however, may be also provided as a precursor which usually has a length of 50-90 nucleotides, particularly 60-80 nucleotides. It should be noted that the precursor may be produced by processing of a primary transcript which may have a length of >100 nucleotides.

The nucleic acid molecules may be present in single-stranded or double-stranded form. The miRNA as such is usually a single-stranded molecule, while the mi-precursor is usually an at least partially self-complementary molecule capable of forming double-stranded portions, e.g. stem- and loop-structures. DNA molecules encoding the miRNA and miRNA precursor molecules. The nucleic acids may be selected from RNA, DNA or nucleic acid analog molecules, such as sugar- or backbone-modified ribonucleotides or deoxyribonucleotides. It should be noted, however, that other nucleic analogs, such as peptide nucleic acids (PNA) or locked nucleic acids (LNA), are also suitable.

In an embodiment of the invention the nucleic acid molecule is an RNA- or DNA molecule, which contains at least one modified nucleotide analog, i.e. a naturally occurring ribonucleotide or deoxyribonucleotide is substituted by a non-naturally occurring nucleotide. The modified nucleotide analog may be located for example at the 5′-end and/or the 3′-end of the nucleic acid molecule.

Preferred nucleotide analogs are selected from sugar- or backbone-modified ribonucleotides. It should be noted, however, that also nucleobase-modified ribonucleotides, i.e. ribonucleotides, containing a non-naturally occurring nucleobase instead of a naturally occurring nucleobase such as uridines or cytidines modified at the 5-position, e.g. 5-(2-amino)propyl uridine, 5-bromo uridine; adenosines and guanosines modified at the 8-position, e.g. 8-bromo guanosine; deaza nucleotides, e.g. 7-deaza-adenosine; O- and N-alkylated nucleotides, e.g. N6-methyl adenosine are suitable. In preferred sugar-modified ribonucleotides the 2′-OH-group is replaced by a group selected from H, OR, R, halo, SH, SR, NH_2, NHR, NR_2 or CN, wherein R is C_1-C_6 alkyl, alkenyl or alkynyl and halo is F, Cl, Br or I. In preferred backbone-modified ribonucleotides the phosphoester group connecting to adjacent ribonucleotides is replaced by a modified group, e.g. of phosphothioate group. It should be noted that the above modifications may be combined.

The nucleic acid molecules of the invention may be obtained by chemical synthesis methods or by recombinant methods, e.g. by enzymatic transcription from synthetic DNA-templates or from DNA-plasmids isolated from recombinant organisms. Typically phage RNA-polymerases are used for transcription, such as T7, T3 or SP6 RNA-polymerases.

The invention also relates to a recombinant expression vector comprising a recombinant nucleic acid operatively linked to an expression control sequence, wherein expression, i.e. transcription and optionally further

processing results in a miRNA-molecule or miRNA precursor molecule as described above. The vector is preferably a DNA-vector, e.g. a viral vector or a plasmid, particularly an expression vector suitable for nucleic acid expression in eukaryotic, more particularly mammalian cells. The recombinant nucleic acid contained in said vector may be a sequence which results in the transcription of the miRNA-molecule as such, a precursor or a primary transcript thereof, which may be further processed to give the miRNA-molecule.

Further, the invention relates to diagnostic or therapeutic applications of the claimed nucleic acid molecules. For example, miRNAs may be detected in biological samples, e.g. in tissue sections, in order to determine and classify certain cell types or tissue types or miRNA-associated pathogenic disorders which are characterized by differential expression of miRNA-molecules or miRNA-molecule patterns. Further, the developmental stage of cells may be classified by determining temporarily expressed miRNA-molecules.

Further, the claimed nucleic acid molecules are suitable for therapeutic applications. For example, the nucleic acid molecules may be used as modulators or targets of developmental processes or disorders associated with developmental dysfunctions, such as cancer. For example, miR-15 and miR-16 probably function as tumor-suppressors and thus expression or delivery of these RNAs or analogs or precursors thereof to tumor cells may provide therapeutic efficacy, particularly against leukemias, such as B-cell chronic lymphocytic leukemia (B-CLL). Further, miR-10 is a possible regulator of the translation of Hox Genes, particularly Hox 3 and Hox 4 (or Scr and Dfd in Drosophila).

In general, the claimed nucleic acid molecules may be used as a modulator of the expression of genes which are at least partially complementary to said nucleic acid. Further, miRNA molecules may act as target for

therapeutic screening procedures, e.g. inhibition or activation of miRNA molecules might modulate a cellular differentiation process, e.g. apoptosis.

Furthermore, existing miRNA molecules may be used as starting materials for the manufacture of sequence-modified miRNA molecules, in order to modify the target-specificity thereof, e.g. an oncogene, a multidrug-resistance gene or another therapeutic target gene. The novel engineered miRNA molecules preferably have an identity of at least 80% to the starting miRNA, e.g. as depicted in Tables 1, 2, 3 and 4. Further, miRNA molecules can be modified, in order that they are symetrically processed and then generated as double-stranded siRNAs which are again directed against therapeutically relevant targets.

Furthermore, miRNA molecules may be used for tissue reprogramming procedures, e.g. a differentiated cell line might be transformed by expression of miRNA molecules into a different cell type or a stem cell.

For diagnostic or therapeutic applications, the claimed RNA molecules are preferably provided as a pharmaceutical composition. This pharmaceutical composition comprises as an active agent at least one nucleic acid molecule as described above and optionally a pharmaceutically acceptable carrier.

The administration of the pharmaceutical composition may be carried out by known methods, wherein a nucleic acid is introduced into a desired target cell in vitro or in vivo.

Commonly used gene transfer techniques include calcium phosphate, DEAE-dextran, electroporation and microinjection and viral methods [30, 31, 32, 33, 34]. A recent addition to this arsenal of techniques for the introduction of DNA into cells is the use of cationic liposomes [35].

Commercially available cationic lipid formulations are e.g. Tfx 50 (Promega) or Lipofectamin 2000 (Life Technologies).

The composition may be in form of a solution, e.g. an injectable solution, a cream, ointment, tablet, suspension or the like. The composition may be administered in any suitable way, e.g. by injection, by oral, topical, nasal, rectal application etc. The carrier may be any suitable pharmaceutical carrier. Preferably, a carrier is used, which is capable of increasing the efficacy of the RNA molecules to enter the target-cells. Suitable examples of such carriers are liposomes, particularly cationic liposomes.

Further, the invention relates to a method of identifying novel microRNA-molecules and precursors thereof, in eukaryotes, particularly in vertebrates and more particularly in mammals, such as humans or mice. This method comprises: ligating 5'- and 3'-adapter-molecules to the end of a size-fractionated RNA-population, reverse transcribing said adapter-ligated RNA-population, and characterizing said reverse transcribed RNA-molecules, e.g. by amplification, concatamerization, cloning and sequencing.

A method as described above already has been described in (8), however, for the identification of siRNA molecules. Surprisingly, it was found now that the method is also suitable for identifying the miRNA molecules or precursors thereof as claimed in the present application.

Further, it should be noted that as 3'-adaptor for derivatization of the 3'-OH group not only 4-hydroxymethylbenzyl but other types of derivatization groups, such as alkyl, alkyl amino, ethylene glycol or 3'-deoxy groups are suitable.

Further, the invention shall be explained in more detail by the following Figures and Examples:

Figure Legends

Fig. 1A. Expression of *D. melanogaster* miRNAs. Northern blots of total RNA isolated from staged populations of *D. melanogaster* were probed for the indicated miRNAs. The position of 76-nt val-tRNA is also indicated on the blots. 5S rRNA serves as loading control. E, embryo; L, larval stage; P, pupae; A, adult; S2, Schneider-2 cells. It should be pointed out, that S2 cells are polyclonal, derived from an unknown subset of embryonic tissues, and may have also lost some features of their tissue of origin while maintained in culture. miR-3 to miR-6 RNAs were not detectable in S2 cells (data not shown). miR-14 was not detected by Northern blotting and may be very weakly expressed, which is consistent with its cloning frequency. Similar miRNA sequences are difficult to distinguish by Northern blotting because of potential cross-hybridization of probes.

Fig. 1B. Expression of vertebrate miRNAs. Northern blots of total RNA isolated from HeLa cells, mouse kidneys, adult zebrafish, frog ovaries, and S2 cells were probed for the indicated miRNAs. The position of 76-nt val-tRNA is also indicated on the blots. 5S rRNA from the preparations of total RNA from the indicated species is also shown. The gels used for probing of miR-18, miR-19a, miR-30, and miR-31 were not run as far as the other gels (see tRNA marker position). miR-32 and miR-33 were not detected by Northern blotting, which is consistent with their low cloning frequency. Oligodeoxynucleotides used as Northern probes were:
let-7a, 5´ TACTATACAACCTACTACCTCAATTTGCC (SEQ ID NO:1);
let-7d, 5´ ACTATGCAACCTACTACCTCT (SEQ ID NO:2);
let-7e, 5´ ACTATACAACCTCCTACCTCA (SEQ ID NO:3);
D. melanogaster val-tRNA, 5´ TGGTGTTTCCGCCCGGGAA (SEQ ID NO:4);
miR-1, 5´ TGGAATGTAAAGAAGTATGGAG (SEQ ID NO:5);
miR-2b, 5´ GCTCCTCAAAGCTGGCTGTGATA (SEQ ID NO:6);
miR-3, 5´ TGAGACACACTTTGCCCAGTGA (SEQ ID NO:7);
miR-4, 5´ TCAATGGTTGTCTAGCTTTAT (SEQ ID NO:8);

miR-5, 5´ CATATCACAACGATCGTTCCTTT (SEQ ID NO:9);
miR-6, 5´ AAAAAGAACAGCCACTGTGATA (SEQ ID NO:10);
miR-7, 5´ TGGAAGACTAGTGATTTTGTTGT (SEQ ID NO:11);
miR-8, 5´ GACATCTTTACCTGACAGTATTA (SEQ ID NO:12);
miR-9, 5´ TCATACAGCTAGATAACCAAAGA (SEQ ID NO:13);
miR-10, 5´ ACAAATTCGGATCTACAGGGT (SEQ ID NO:14);
miR-11, 5´ GCAAGAACTCAGACTGTGATG (SEQ ID NO:15);
miR-12, 5´ ACCAGTACCTGATGTAATACTCA (SEQ ID NO:16);
miR-13a, 5´ ACTCGTCAAAATGGCTGTGATA (SEQ ID NO:17);
miR-14, 5' TAGGAGAGAGAAAAAGACTGA (SEQ ID NO:18);
miR-15, 5´ TAGCAGCACATAATGGTTTGT (SEQ ID NO:19);
miR-16, 5´ GCCAATATTTACGTGCTGCTA (SEQ ID NO:20);
miR-17, 5´ TACAAGTGCCTTCACTGCAGTA (SEQ ID NO:21);
miR-18, 5´ TATCTGCACTAGATGCACCTTA (SEQ ID NO:22);
miR-19a, 5´ TCAGTTTTGCATAGATTTGCACA (SEQ ID NO:23);
miR-20, 5´ TACCTGCACTATAAGCACTTTA (SEQ ID NO:24);
miR-21, 5´ TCAACATCAGTCTGATAAGCTA (SEQ ID NO:25);
miR-22, 5´ ACAGTTCTTCAACTGGCAGCTT (SEQ ID NO:26);
miR-23, 5´ GGAAATCCCTGGCAATGTGAT (SEQ ID NO:27);
miR-24, 5´ CTGTTCCTGCTGAACTGAGCCA (SEQ ID NO:28);
miR-25, 5´ TCAGACCGAGACAAGTGCAATG (SEQ ID NO:29);
miR-26a, 5´ AGCCTATCCTGGATTACTTGAA (SEQ ID NO:30);
miR-27; 5´ AGCGGAACTTAGCCACTGTGAA (SEQ ID NO:31);
miR-28, 5´ CTCAATAGACTGTGAGCTCCTT (SEQ ID NO:32);
miR-29, 5´ AACCGATTTCAGATGGTGCTAG (SEQ ID NO:33);
miR-30, 5´ GCTGCAAACATCCGACTGAAAG (SEQ ID NO:34);
miR-31, 5´ CAGCTATGCCAGCATCTTGCCT (SEQ ID NO:35);
miR-32, 5' GCAACTTAGTAATGTGCAATA (SEQ ID NO:36);
miR-33, 5' TGCAATGCAACTACAATGCACC (SEQ ID NO:37).

Fig. 2. Genomic organization of miRNA gene clusters. The precursor structure is indicated as box and the location of the miRNA within the

precursor is shown in gray; the chromosomal location is also indicated to the right. (A) D. melanogaster miRNA gene clusters. (B) Human miRNA gene clusters. The cluster of let-7a-1 and let-7f-1 is separated by 26500 nt from a copy of let-7d on chromosome 9 and 17. A cluster of let-7a-3 and let-7b, separated by 938 nt on chromosome 22, is not illustrated.

Fig. 3. Predicted precursor structures of D. melanogaster miRNAs. RNA secondary structure prediction was performed using mfold version 3.1 [28] and manually refined to accommodate G/U wobble base pairs in the helical segments. The miRNA sequence is underlined. The actual size of the stem-loop structure is not known experimentally and may be slightly shorter or longer than represented. Multicopy miRNAs and their corresponding precursor structures are also shown.

Fig. 4. Predicted precursor structures of human miRNAs. For legend, see Fig. 3.

Fig. 5. Expression of novel mouse miRNAs. Northern blot analysis of novel mouse miRNAs. Total RNA from different mouse tissues was blotted and probed with a 5´-radiolabeled oligodeoxynucleotide complementary to the indicated miRNA. Equal loading of total RNA on the gel was verified by ethidium bromide staining prior to transfer; the band representing tRNAs is shown. The fold-back precursors are indicated with capital L. Mouse brains were dissected into midbrain, mb, cortex, cx, cerebellum, cb. The rest of the brain, rb, was also used. Other tissues were heart, ht, lung, lg, liver, lv, colon, co, small intestine, si, pancreas, pc, spleen, sp, kidney, kd, skeletal muscle, sm, stomach, st, H, human Hela SS3 cells. Oligodeoxynucleotides used as Northern probes were:
miR-1a, CTCCATACTTCTTTACATTCCA (SEQ ID NO:38);
miR-30b, GCTGAGTGTAGGATGTTTACA (SEQ ID NO:39);
miR-30a-s, GCTTCCAGTCGAGGATGTTTACA (SEQ ID NO:40);
miR-99b, CGCAAGGTCGGTTCTACGGGTG (SEQ ID NO:41);

miR-101, TCAGTTATCACAGTACTGTA (SEQ ID NO:42);
miR-122a, ACAAACACCATTGTCACACTCCA (SEQ ID NO:43);
miR-124a, TGGCATTCACCGCGTGCCTTA (SEQ ID NO:44);
miR-125a, CACAGGTTAAAGGGTCTCAGGGA (SEQ ID NO:45);
miR-125b, TCACAAGTTAGGGTCTCAGGGA (SEQ ID NO:46);
miR-127, AGCCAAGCTCAGACGGATCCGA (SEQ ID NO:47);
miR-128, AAAAGAGACCGGTTCACTCTGA (SEQ ID NO:48);
miR-129, GCAAGCCCAGACCGAAAAAAG (SEQ ID NO:49);
miR-130, GCCCTTTTAACATTGCACTC (SEQ ID NO:50);
miR-131, ACTTTCGGTTATCTAGCTTTA (SEQ ID NO:51);
miR-132, ACGACCATGGCTGTAGACTGTTA (SEQ ID NO:52);
miR-143, TGAGCTACAGTGCTTCATCTCA (SEQ ID NO:53).

Fig.6. Potential orthologs of lin-4 stRNA. (A) Sequence alignment of *C. elegans* lin-4 stRNA with mouse miR-125a and miR-125b and the *D. melanogaster* miR-125. Differences are highlighted by gray boxes. (B) Northern blot of total RNA isolated from staged populations of *D. melanogaster*, probed for miR-125. E, embryo; L, larval stage; P, pupae; A, adult; S2, Schneider-2 cells.

Fig. 7. Predicted precursor structures of miRNAs, sequence accession numbers and homology information. RNA secondary structure prediction was performed using mfold version 3.1 and manually refined to accommodate G/U wobble base pairs in the helical segments. Dashes were inserted into the secondary structure presentation when asymmetrically bulged nucleotides had to be accommodated. The excised miRNA sequence is underlined. The actual size of the stem-loop structure is not known experimentally and may be slightly shorter or longer than represented. Multicopy miRNAs and their corresponding precursor structures are also shown. In cases where no mouse precursors were yet deposited in the database, the human orthologs are indicated. miRNAs

which correspond to *D. melanogaster* or human sequences are included. Published *C. elegans* miRNAs [36, 37] are also included in the table. A recent set of new HeLa cell miRNAs is also indicated [46]. If several ESTs were retrieved for one organism in the database, only those with different precursor sequences are listed. miRNA homologs found in other species are indicated. Chromosomal location and sequence accession numbers, and clusters of miRNA genes are indicated. Sequences from cloned miRNAs were searched against mouse and human in GenBank (including trace data), and against *Fugu rubripes* and *Danio rerio* at www.jgi.doe.gov and www.sanger.ac.uk, respectively.

EXAMPLE 1: MicroRNAs from D. melanogaster and human.

We previously developed a directional cloning procedure to isolate siRNAs after processing of long dsRNAs in Drosophila melanogaster embryo lysate (8). Briefly, 5´ and 3´ adapter molecules were ligated to the ends of a size-fractionated RNA population, followed by reverse transcription, PCR amplification, concatamerization, cloning and sequencing. This method, originally intended to isolate siRNAs, led to the simultaneous identification of 14 novel 20- to 23-nt short RNAs which are encoded in the D. melanogaster genome and which are expressed in 0 to 2 h embryos (Table 1). The method was adapted to clone RNAs in a similar size range from HeLa cell total RNA (14), which led to the identification of 19 novel human stRNAs (Table 2), thus providing further evidence for the existence of a large class of small RNAs with potential regulatory roles. According to their small size, we refer to these novel RNAs as microRNAs or miRNAs. The miRNAs are abbreviated as miR-1 to miR-33, and the genes encoding miRNAs are named mir-1 to mir-33. Highly homologous miRNAs are classified by adding a lowercase letter, followed by a dash and a number for designating multiple genomic copies of a mir gene.

The expression and size of the cloned, endogenous short RNAs was also examined by Northern blotting (Fig. 1, Table 1 and 2). Total RNA isolation was performed by acid guanidinium thiocyanate-phenol-chloroform extraction [45]. Northern analysis was performed as described [1], except that the total RNA was resolved on a 15% denaturing polyacrylamide gel, transferred onto Hybond-N+membrane (Amersham Pharmacia Biotech), and the hybridization and wash steps were performed at 50°C. Oligodeoxynucleotides used as Northern probes were 5'-32P-phosphorylated, complementary to the miRNA sequence and 20 to 25 nt in length.

5S rRNA was detected by ethidium staining of polyacrylamide gels prior to transfer. Blots were stripped by boiling in 0.1% aqueous sodium dodecylsulfate/0.1x SSC (15 mM sodium chloride, 1.5 mM sodium citrate, pH 7.0) for 10 min, and were re-probed up to 4 times until the 21-nt signals became too weak for detection. Finally, blots were probed for val-tRNA as size marker.

For analysis of D. melanogaster RNAs, total RNA was prepared from different developmental stages, as well as cultured Schneider-2 (S2) cells, which originally derive from 20-24 h D. melanogaster embryos [15] (Fig. 1, Table 1). miR-3 to miR-7 are expressed only during embryogenesis and not at later developmental stages. The temporal expression of miR-1, miR-2 and miR-8 to miR-13 was less restricted. These miRNAs were observed at all developmental stages though significant variations in the expression levels were sometimes observed. Interestingly, miR-1, miR-3 to miR-6, and miR-8 to miR-11 were completely absent from cultured Schneider-2 (S2) cells, which were originally derived from 20-24 h D. melanogaster embryos [15], while miR-2, miR-7, miR-12, and miR-13 were present in S2 cells, therefore indicating cell type-specific miRNA expression. miR-1, miR-8, and miR-12 expression patterns are similar to those of lin-4 stRNA in C. elegans, as their expression is strongly upregulated in larvae and sustained

to adulthood [16]. miR-9 and miR-11 are present at all stages but are strongly reduced in the adult which may reflect a maternal contribution from germ cells or expression in one sex only.

The mir-3 to mir-6 genes are clustered (Fig. 2A), and mir-6 is present as triple repeat with slight variations in the mir-6 precursor sequence but not in the miRNA sequence itself. The expression profiles of miR-3 to miR-6 are highly similar (Table 1), which suggests that a single embryo-specific precursor transcript may give rise to the different miRNAs, or that the same enhancer regulates miRNA-specific promoters. Several other fly miRNAs are also found in gene clusters (Fig. 2A).

The expression of HeLa cell miR-15 to miR-33 was examined by Northern blotting using HeLa cell total RNA, in addition to total RNA prepared from mouse kidneys, adult zebrafish, Xenopus laevis ovary, and D. melanogaster S2 cells (Fig. 1B, Table 2). miR-15 and miR-16 are encoded in a gene cluster (Fig. 2B) and are detected in mouse kidney, fish, and very weakly in frog ovary, which may result from miRNA expression in somatic ovary tissue rather than oocytes. mir-17 to mir-20 are also clustered (Fig. 2B), and are expressed in HeLa cells and fish, but undetectable in mouse kidney and frog ovary (Fig. 1, Table 2), and therefore represent a likely case of tissue-specific miRNA expression.

The majority of vertebrate and invertebrate miRNAs identified in this study are not related by sequence, but a few exceptions, similar to the highly conserved let-7 RNA [6], do exist. Sequence analysis of the D. melanogaster miRNAs revealed four such examples of sequence conservation between invertebrates and vertebrates. miR-1 homologs are encoded in the genomes of C. elegans, C. briggsae, and humans, and are found in cDNAs from zebrafish, mouse, cow and human. The expression of mir-1 was detected by Northern blotting in total RNA from adult zebrafish and C. elegans, but not in total RNA from HeLa cells or mouse kidney

(Table 2 and data not shown). Interestingly, while mir-1 and let-7 are expressed both in adult flies (Fig. 1A) [6] and are both undetected in S2 cells, miR-1 is, in contrast to let-7, undetectable in HeLa cells. This represents another case of tissue-specific expression of a miRNA, and indicates that miRNAs may not only play a regulatory role in developmental timing, but also in tissue specification. miR-7 homologs were found by database searches in mouse and human genomic and expressed sequence tag sequences (ESTs). Two mammalian miR-7 variants are predicted by sequence analysis in mouse and human, and were detected by Northern blotting in HeLa cells and fish, but not in mouse kidney (Table 2). Similarly, we identified mouse and human miR-9 and miR-10 homologs by database searches but only detected mir-10 expression in mouse kidney.

The identification of evolutionary related miRNAs, which have already acquired multiple sequence mutations, was not possible by standard bioinformatic searches. Direct comparison of the D. melanogaster miRNAs with the human miRNAs identified an 11-nt segment shared between D. melanogaster miR-6 and HeLa miR-27, but no further relationships were detected. One may speculate that most miRNAs only act on a single target and therefore allow for rapid evolution by covariation, and that highly conserved miRNAs act on more than one target sequence, and therefore have a reduced probability for evolutionary drift by covariation [6]. An alternative interpretation is that the sets of miRNAs from D. melanogaster and humans are fairly incomplete and that many more miRNAs remain to be discovered, which will provide the missing evolutionary links.

lin-4 and let-7 stRNAs were predicted to be excised from longer transcripts that contain approximately 30 base-pair stem-loop structures [1, 6]. Database searches for newly identified miRNAs revealed that all miRNAs are flanked by sequences that have the potential to form stable stem-loop structures (Fig. 3 and 4). In many cases, we were able to detect the predicted, approximately 70-nt precursors by Northern blotting (Fig. 1).

Some miRNA precursor sequences were also identified in mammalian cDNA (EST) databases [27], indicating that primary transcripts longer than 70-nt stem-loop precursors do also exist. We never cloned a 22-nt RNA complementary to any of the newly identified miRNAs, and it is as yet unknown how the cellular processing machinery distinguishes between the miRNA and its complementary strand. Comparative analysis of the precursor stem-loop structures indicates that the loops adjacent to the base-paired miRNA segment can be located on either side of the miRNA sequence (Fig. 3 and 4), suggesting that the 5′ or 3′ location of the stem-closing loop is not the determinant of miRNA excision. It is also unlikely that the structure, length or stability of the precursor stem is the critical determinant as the base-paired structures are frequently imperfect and interspersed by less stable, non-Watson-Crick base pairs such as G/A, U/U, C/U, A/A, and G/U wobbles. Therefore, a sequence-specific recognition process is a likely determinant for miRNA excision, perhaps mediated by members of the Argonaute (rde-1/ago1/piwi) protein family. Two members of this family, alg-1 and alg-2, have recently been shown to be critical for stRNA processing in C. elegans [13]. Members of the Argonaute protein family are also involved in RNAi and PTGS. In D. melanogaster, these include argonaute2, a component of the siRNA-endonuclease complex (RISC) [17], and its relative aubergine, which is important for silencing of repeat genes [18]. In other species, these include rde-1, argonaute1, and qde-2, in C. elegans [19], Arabidopsis thaliana [20], and Neurospora crassa [21], respectively. The Argonaute protein family therefore represents, besides the RNase III Dicer [12, 13], another evolutionary link between RNAi and miRNA maturation.

Despite advanced genome projects, computer-assisted detection of genes encoding functional RNAs remains problematic [22]. Cloning of expressed, short functional RNAs, similar to EST approaches (RNomics), is a powerful alternative and probably the most efficient method for identification of such novel gene products [23-26]. The number of functional RNAs has been

widely underestimated and is expected to grow rapidly because of the development of new functional RNA cloning methodologies.

The challenge for the future is to define the function and the potential targets of these novel miRNAs by using bioinformatics as well as genetics, and to establish a complete catalogue of time- and tissue-specific distribution of the already identified and yet to be uncovered miRNAs. lin-4 and let-7 stRNAs negatively regulate the expression of proteins encoded by mRNAs whose 3′ untranslated regions contain sites of complementarity to the stRNA [3-5].

Thus, a series of 33 novel genes, coding for 19- to 23-nucleotide microRNAs (miRNAs), has been cloned from fly embryos and human cells. Some of these miRNAs are highly conserved between vertebrates and invertebrates and are developmentally or tissue-specifically expressed. Two of the characterized human miRNAs may function as tumor suppressors in B-cell chronic lymphocytic leukemia. miRNAs are related to a small class of previously described 21- and 22-nt RNAs (lin-4 and let-7 RNAs), so-called small temporal RNAs (stRNAs), and regulate developmental timing in C. elegans and other species. Similar to stRNAs, miRNAs are presumed to regulate translation of specific target mRNAs by binding to partially complementary sites, which are present in their 3′-untranslated regions.

Deregulation of miRNA expression may be a cause of human disease, and detection of expression of miRNAs may become useful as a diagnostic. Regulated expression of miRNAs in cells or tissue devoid of particular miRNAs may be useful for tissue engineering, and delivery or transgenic expression of miRNAs may be useful for therapeutic intervention. miRNAs may also represent valuable drug targets itself. Finally, miRNAs and their precursor sequences may be engineered to recognize therapeutic valuable targets.

EXAMPLE 2: miRNAs from mouse.

To gain more detailed insights into the distribution and function of miRNAs in mammals, we investigated the tissue-specific distribution of miRNAs in adult mouse. Cloning of miRNAs from specific tissues was preferred over whole organism-based cloning because low-abundance miRNAs that normally go undetected by Northern blot analysis are identified clonally. Also, in situ hybridization techniques for detecting 21-nt RNAs have not yet been developed. Therefore, 19- to 25-nucleotide RNAs were cloned and sequenced from total RNA, which was isolated from 18.5 weeks old BL6 mice. Cloning of miRNAs was performed as follows: 0.2 to 1 mg of total RNA was separated on a 15% denaturing polyacrylamide gel and RNA of 19- to 25-nt size was recovered. A 5′-phosphorylated 3′-adapter oligonucleotide (5′-pUUUaaccgcgaattccagx: uppercase, RNA; lowercase, DNA; p, phosphate; x, 3′-Amino-Modifier C-7, ChemGenes, Ashland, Ma, USA, Cat. No. NSS-1004; SEQ ID NO:54) and a 5′-adapter oligonucleotide (5′-acggaattcctcactAAA: uppercase, RNA; lowercase, DNA; SEQ ID NO:55) were ligated to the short RNAs. RT/PCR was performed with 3′-primer (5′-GACTAGCTGGAATTCGCGGTTAAA; SEQ ID NO:56) and 5′-primer (5′-CAGCCAACGGAATTCCTCACTAAA; SEQ ID NO:57). In order to introduce Ban I restriction sites, a second PCR was performed using the primer pair 5′-CAGCCAACAGGCACCGAATTCCTCACTAAA (SEQ ID NO:57) and 5′-GACTAGCTTGGTGCCGAATTCGCGGTTAAA (SEQ ID NO:56), followed by concatamerization after Ban I digestion and T4 DNA ligation. Concatamers of 400 to 600 basepairs were cut out from 1.5% agarose gels and recovered by Biotrap (Schleicher & Schuell) electroelution (1x TAE buffer) and by ethanol precipitation. Subsequently, the 3′ ends of the concatamers were filled in by incubating for 15 min at 72°C with Taq polymerase in standard PCR reaction mixture. This solution was diluted 3-fold with water and directly used for ligation into pCR2.1 TOPO vectors. Clones were screened for inserts by PCR and 30 to 50 samples were subjected to sequencing. Because RNA was prepared from combining

tissues of several mice, minor sequence variations that were detected multiple times in multiple clones may reflect polymorphisms rather than RT/PCR mutations. Public database searching was used to identify the genomic sequences encoding the approx. 21-nt RNAs. The occurrence of a 20 to 30 basepair fold-back structure involving the immediate upstream or downstream flanking sequences was used to assign miRNAs [36-38].

We examined 9 different mouse tissues and identified 34 novel miRNAs, some of which are highly tissue-specifically expressed (Table 3 and Figure 5). Furthermore, we identified 33 new miRNAs from different mouse tissues and also from human Soas-2 osteosarcoma cells (Table 4). miR-1 was previously shown by Northern analysis to be strongly expressed in adult heart, but not in brain, liver, kidney, lung or colon [37]. Here we show that miR-1 accounts for 45% of all mouse miRNAs found in heart, yet miR-1 was still expressed at a low level in liver and midbrain even though it remained undetectable by Northern analysis. Three copies or polymorphic alleles of miR-1 were found in mice. The conservation of tissue-specific miR-1 expression between mouse and human provides additional evidence for a conserved regulatory role of this miRNA. In liver, variants of miR-122 account for 72% of all cloned miRNAs and miR-122 was undetected in all other tissues analyzed. In spleen, miR-143 appeared to be most abundant, at a frequency of approx. 30%. In colon, miR-142-as, was cloned several times and also appeared at a frequency of 30%. In small intestine, too few miRNA sequences were obtained to permit statistical analysis. This was due to strong RNase activity in this tissue, which caused significant breakdown of abundant non-coding RNAs, e.g. rRNA, so that the fraction of miRNA in the cloned sequences was very low. For the same reason, no miRNA sequences were obtained from pancreas.

To gain insights in neural tissue miRNA distribution, we analyzed cortex, cerebellum and midbrain. Similar to heart, liver and small intestine, variants

of a particular miRNA, miR-124, dominated and accounted for 25 to 48% of all brain miRNAs. miR-101, -127, -128, -131, and -132, also cloned from brain tissues, were further analyzed by Northern blotting and shown to be predominantly brain-specific. Northern blot analysis was performed as described in Example 1. tRNAs and 5S rRNA were detected by ethidium staining of polyacrylamide gels prior to transfer to verify equal loading. Blots were stripped by boiling in deionized water for 5 min, and reprobed up to 4 times until the 21-nt signals became too weak for detection.

miR-125a and miR-125b are very similar to the sequence of *C. elegans* lin-4 stRNA and may represent its orthologs (Fig. 6A). This is of great interest because, unlike let-7 that was readily detected in other species, lin-4 has acquired a few mutations in the central region and thus escaped bioinformatic database searches. Using the mouse sequence miR-125b, we could readily identify its ortholog in the *D. melanogaster* genome. miR-125a and miR-125b differ only by a central diuridine insertion and a U to C change. miR-125b is very similar to lin-4 stRNA with the differences located only in the central region, which is presumed to be bulged out during target mRNA recognition [41]. miR-125a and miR-125b were cloned from brain tissue, but expression was also detected by Northern analysis in other tissues, consistent with the role for lin-4 in regulating neuronal remodeling by controlling lin-14 expression [43]. Unfortunately, orthologs to *C. elegans* lin-14 have not been described and miR-125 targets remain to be identified in *D. melanogaster* or mammals. Finally, miR-125b expression is also developmentally regulated and only detectable in pupae and adult but not in embryo or larvae of *D. melanogaster* (Fig. 6B).

Sequence comparison of mouse miRNAs with previously described miRNA reveals that miR-99b and miR-99a are similar to *D. melanogaster*, mouse and human miR-10 as well as *C. elegans* miR-51 [36], miR-141 is similar to *D. melanogaster* miR-8 , miR-29b is similar to *C. elegans* miR-83 , and miR-131 and miR-142-s are similar to *D. melanogaster* miR-4 and *C.*

elegans miR-79 [36]. miR-124a is conserved between invertebrates and vertebrates. In this respect it should be noted that for almost every miRNA cloned from mouse was also encoded in the human genome, and frequently detected in other vertebrates, such as the pufferfish, *Fugu rubripes*, and the zebrafish, *Danio rerio*. Sequence conservation may point to conservation in function of these miRNAs. Comprehensive information about orthologous sequences is listed in Fig. 7.

In two cases both strands of miRNA precursors were cloned (Table 3), which was previously observed once for a *C. elegans* miRNA [36]. It is thought that the most frequently cloned strand of a miRNA precursor represents the functional miRNA, which is miR-30c-s and miR-142-as, s and as indicating the 5´ or 3´ side of the fold-back structure, respectively.

The mir-142 gene is located on chromosome 17, but was also found at the breakpoint junction of a t(8;17) translocation, which causes an aggressive B-cell leukemia due to strong up-regulation of a translocated MYC gene [44]. The translocated MYC gene, which was also truncated at the first exon, was located only 4-nt downstream of the 3´-end of the miR-142 precursor. This suggests that translocated MYC was under the control of the upstream miR-142 promoter. Alignment of mouse and human miR-142 containing EST sequences indicate an approximately 20 nt conserved sequence element downstream of the mir-142 hairpin. This element was lost in the translocation. It is conceivable that the absence of the conserved downstream sequence element in the putative miR-142/mRNA fusion prevented the recognition of the transcript as a miRNA precursor and therefore may have caused accumulation of fusion transcripts and overexpression of MYC.

miR-155, which was cloned from colon, is excised from the known noncoding BIC RNA [47]. BIC was originally identified as a gene transcriptionally activated by promoter insertion at a common retroviral

integration site in B cell lymphomas induced by avian leukosis virus. Comparison of BIC cDNAs from human, mouse and chicken revealed 78% identity over 138 nucleotides [47]. The identity region covers the miR-155 fold-back precursor and a few conserved boxes downstream of the fold-back sequence. The relatively high level of expression of BIC in lymphoid organs and cells in human, mouse and chicken implies an evolutionary conserved function, but BIC RNA has also been detected at low levels in non-hematopoietic tissues [47].

Another interesting observation was that segments of perfect complementarity to miRNAs are not observed in mRNA sequences or in genomic sequences outside the miRNA inverted repeat. Although this could be fortuitous, based on the link between RNAi and miRNA processing [11, 13, 43] it may be speculated that miRNAs retain the potential to cleave perfectly complementary target RNAs. Because translational control without target degradation could provide more flexibility it may be preferred over mRNA degradation.

In summary, 63 novel miRNAs were identified from mouse and 4 novel miRNAs were identified from human Soas-2 osteosarcoma cells (Table 3 and Table 4), which are conserved in human and often also in other non-mammalian vertebrates. A few of these miRNAs appear to be extremely tissue-specific, suggesting a critical role for some miRNAs in tissue-specification and cell lineage decisions. We may have also identified the fruitfly and mammalian ortholog of *C. elegans* lin-4 stRNA. The establishment of a comprehensive list of miRNA sequences will be instrumental for bioinformatic approaches that make use of completed genomes and the power of phylogenetic comparison in order to identify miRNA-regulated target mRNAs.

References and Notes

1. R. C. Lee, R. L. Feinbaum, V. Ambros, Cell 75, 843 (1993).
2. B. J. Reinhart et al., Nature 403, 901 (2000).
3. V. Ambros, Curr. Opin. Genet. Dev. 10, 428 (2000).
4. E. G. Moss, Curr. Biol. 10, R436 (2000).
5. F. Slack, G. Ruvkun, Annu. Rev. Genet. 31, 611 (1997).
6. A. E. Pasquinelli et al., Nature 408, 86 (2000).
7. S. M. Elbashir et al., Nature 411, 494 (2001).
8. S. M. Elbashir, W. Lendeckel, T. Tuschl, Genes & Dev. 15, 188 (2001).
9. A. J. Hamilton, D. C. Baulcombe, Science 286, 950 (1999).
10. S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, Nature 404, 293 (2000).
11. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, Cell 101, 25 (2000).
12. G. Hutvágner, J. McLachlan, É. Bálint, T. Tuschl, P. D. Zamore, Science 93, 834 (2001).
13. A. Grishok et al., Cell 106, 23 (2001).
14. Cloning of 19- to 24-nt RNAs from D. melanogaster 0-2 h embryo lysate was performed as described (8). For cloning of HeLa miRNAs, 1 mg of HeLa total RNA was separated on a 15% denaturing polyacrylamide gel and RNA of 19- to 25-nt size was recovered. A 5′ phosphorylated 3′ adapter oligonucleotide (5′ pUUU-aaccgcgaattccagx: uppercase, RNA; lowercase, DNA; p, phosphate; x, 4-hydroxymethylbenzyl; SEQ ID NO:54) and a 5′ adapter oligonucleotide (5′ acggaattcctcactAAA: uppercase, RNA; lowercase, DNA; SEQ ID NO:55) were ligated to the short HeLa cell RNAs. RT/PCR was performed with 3′ primer (5′ GACTAGCTGGAATTCGCGGTTAAA; SEQ ID NO:56) and 5′ primer (5′ CAGCCAACGGAATTCCTCACTAAA; SEQ ID NO:57), and followed by concatamerization after Eco RI digestion and T4 DNA

ligation (8). After ligation of concatamers into pCR2.1 TOPO vectors, about 100 clones were selected and subjected to sequencing.

15. I. Schneider, J Embryol Exp Morphol 27, 353 (1972).
16. R. Feinbaum, V. Ambros, Dev. Biol. 210, 87 (1999).
17. S. M. Hammond, S. Boettcher, A. A. Caudy, R. Kobayashi, G. J. Hannon, Science 293, 1146 (2001).
18. A. A. Aravin et al., Curr. Biol. 11, 1017 (2001).
19. H. Tabara et al., Cell 99, 123 (1999).
20. M. Fagard, S. Boutet, J. B. Morel, C. Bellini, H. Vaucheret, Proc. Natl. Acad. Sci. USA 97, 11650 (2000).
21. C. Catalanotto, G. Azzalin, G. Macino, C. Cogoni, Nature 404, 245 (2000).
22. S. R. Eddy, Curr. Opin. Genet. Dev. 9, 695 (1999).
23. J. Cavaille et al., Proc. Natl. Acad. Sci. USA 97, 14311 (2000).
24. A. Hüttenhofer et al., EMBO J. 20, 2943 (2001).
25. L. Argaman et al., Curr. Biol. 11, 941 (2001).
26. K. M. Wassarman, F. Repoila, C. Rosenow, G. Storz, S. Gottesman, Genes & Dev. 15, 1637 (2001).
27. Supplementary Web material is available on Science Online at www.sciencemag.org/cgi/content/full/xxx
28. D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, J. Mol. Biol. 288, 911 (1999).
29. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, Nature 409, 363 (2001).
30. Graham, F.L. and van der Eb, A.J., (1973), Virol. 52, 456.
31. McCutchan, J.H. and Pagano, J.S., (1968), J. Natl. Cancer Inst. 41, 351.
32. Chu, G. et al., (1987), Nucl. Acids Res. 15, 1311.
33. Fraley, R. et al., (1980), J. Biol. Chem. 255, 10431.
34. Capecchi, M.R., (1980), Cell 22, 479.
35. Felgner, P.L. et al., (1987), Proc. Natl. Acad. Sci USA 84, 7413.

36. Lau N.C., Lim L.P., Weinstein E.G., Bartel D.P., (2001), Science *294*, 858-862.

37. Lee R.C., Ambros V., (2001), Science *294*, 862-864.

38. Ambros V., (2001), Cell *107*, 823-826.

39. Ambros V., Horvitz H.R., (1984), Science *226*, 409-416.

40. Wightman B., Ha I., Ruvkun G., (1993), Cell *75*, 855-862.

41. Rougvie A.E., (2001), Nat. Rev. Genet. *2*, 690-701.

42. Ketting R.F., Fischer S.E., Bernstein E., Sijen T., Hannon G.J., Plasterk R.H., (2001), Genes & Dev. *15*, 2654-2659.

43. Hallam S.J., Jin Y., (1998), Nature *395*, 78-82.

44. Gauwerky C.E., Huebner K., Isobe M., Nowell P.C., Croce C.M., (1989), Proc. Natl. Acad. Sci. USA *86*, 8867-8871.

45. P. Chomczynski, N. Sacchi, Anal Biochem 162, 156, (1987).

46. Mourelatos Z., Dostie J., Paushkin S., Sharma A., Charroux B., Abel L., J.R., Mann M., Dreyfuss G., (2002), Genes & Dev., in press.

47. Tam W., (2001), Gene *274*, 157-167.

Table 1

D. melanogaster miRNAs. The sequences given represent the most abundant, and typically longest miRNA sequence identified by cloning; miRNAs frequently vary in length by one or two nucleotides at their 3′ termini. From 222 short RNAs sequenced, 69 (31%) corresponded to miRNAs, 103 (46%) to already characterized functional RNAs (rRNA, 7SL RNA, tRNAs), 30 (14%) to transposon RNA fragments, and 20 (10%) sequences with no database entry. The frequency (freq.) for cloning a particular miRNA relative to all identified miRNAs is indicated in percent. Results of Northern blotting of total RNA isolated from staged populations of D. melanogaster are summarized. E, embryo; L, larval stage; P, pupae; A, adult; S2, Schneider-2 cells. The strength of the signal within each blot is represented from strongest (+ + +) to undetected (-). let-7 stRNA was probed as control. Genbank accession numbers and homologs of miRNAs identified by database searching in other species are provided as supplementary material.

miRNA	sequence (5′ to 3′)	freq. (%)	E 0-3 h	E 0-6 h	L1+ L2	L3	P	A	S2
miR-1	UGGAAUGUAAAGAAGUAUGGAG (SEQ ID NO:58)	32	+	+	+++	+++	++	+++	-
miR-2a*	UAUCACAGCCAGCUUUGAUGAGC (SEQ ID NO:59)	3							
miR-2b*	UAUCACAGCCAGCUUUGAGGAGC (SEQ ID NO:60)	3	++	++	++	+++	++	+	+++
miR-3	UCACUGGGCAAAGUGUGUCUCA#	9	+++	+++	-	-	-	-	-
miR-4	AUAAAGCUAGACAACCAUUGA (SEQ ID NO:62)	6	+++	+++	-	-	-	-	-
miR-5	AAAGGAACGAUCGUUGUGAUAUG (SEQ ID NO:63)	1	+++	+++	+/-	+/-	-	-	-
miR-6	UAUCACAGUGGCUGUUCUUUUU (SEQ ID NO:64)	13	+++	+++	+/-	+/-	-	-	-
miR-7	UGGAAGACUAGUGAUUUUGUUGU (SEQ ID NO:65)	4	+++	++	+/-	+/-	+/-	+/-	+/-
miR-8	UAAUACUGUCAGGUAAAGAUGUC (SEQ ID NO:66)	3	+/-	+/-	+++	+++	+	+++	-

miR-9	UCUUUGGUUAUCUAGCUGUAUGA (SEQ ID NO:67)	7	+++	++	+++	+++	+++	+/-	-
miR-10	ACCCUGUAGAUCCGAAUUUGU (SEQ ID NO:68)	1	+	+	++	+++	+/-	+	-
miR-11	CAUCACAGUCUGAGUUCUUGC (SEQ ID NO:69)	7	+++	+++	+++	+++	+++	+	-
miR-12	UGAGUAUUACAUCAGGUACUGGU (SEQ ID NO:70)	7	+	+	++	++	+	+++	+/-
miR-13a*	UAUCACAGCCAUUUUGACGAGU (SEQ ID NO:71)	1	+++	+++	+++	+++	+	+++	+++
miR-13b*	UAUCACAGCCAUUUUGAUGAGU (SEQ ID NO:72)	0							
miR-14	UCAGUCUUUUUCUCUCUCCUA (SEQ ID NO:73)	1	-	-	-	-	-	-	-
let-7	UGAGGUAGUAGGUUGUAUAGUU (SEQ ID NO:74)	0	-	-	-	-	+++	+++	-

= (SEQ ID NO:61)

*Similar miRNA sequences are difficult to distinguish by Northern blotting because of potential cross-hybridization of probes.

Table 2

Human miRNAs. From 220 short RNAs sequenced, 100 (45%) corresponded to miRNAs, 53 (24%) to already characterized functional RNAs (rRNA, snRNAs, tRNAs), and 67 (30%) sequences with no database entry. Results of Northern blotting of total RNA isolated from different vertebrate species and S2 cells are indicated. For legend, see Table 1.

miRNA	sequence (5' to 3')	freq. (%)	HeLa cells	mouse kidney	adult fish	frog ovary	S2
let-7a*	UGAGGUAGUAGGUUGUAUAGUU#	10	+++	+++	+++	-	-
let-7b*	UGAGGUAGUAGGUUGUGUGGUU (SEQ ID NO:76)	13					
let-7c*	UGAGGUAGUAGGUUGUAUGGUU (SEQ ID NO:77)	3					
let-7d*	AGAGGUAGUAGGUUGCAUAGU (SEQ ID NO:78)	2	+++	+++	+++	-	-
let-7e*	UGAGGUAGGAGGUUGUAUAGU (SEQ ID NO:79)	2	+++	+++	+++	-	-
let-7f*	UGAGGUAGUAGAUUGUAUAGUU (SEQ ID NO:80)	1					
miR-15	UAGCAGCACAUAAUGGUUUGUG (SEQ ID NO:81)	3	+++	++	+	+/-	-
miR-16	UAGCAGCACGUAAAUAUUGGCG (SEQ ID NO:82)	10	+++	+	+/-	+/-	-
miR-17	ACUGCAGUGAAGGCACUUGU (SEQ ID NO:83)	1	+++	-	-	-	-
miR-18	UAAGGUGCAUCUAGUGCAGAUA (SEQ ID NO:84)	2	+++	-	-	-	-
miR-19a*	UGUGCAAAUCUAUGCAAAACUGA (SEQ ID NO:85)	1	+++	-	+/-	-	-
miR-19b*	UGUGCAAAUCCAUGCAAAACUGA (SEQ ID NO:86)	3					
miR-20	UAAAGUGCUUAUAGUGCAGGUA (SEQ ID NO:87)	4	+++	-	+	-	-
miR-21	UAGCUUAUCAGACUGAUGUUGA (SEQ ID NO:88)	10	+++	+	++	-	-
miR-22	AAGCUGCCAGUUGAAGAACUGU (SEQ ID NO:89)	10	+++	+++	+	+/-	-
miR-23	AUCACAUUGCCAGGGAUUUCC (SEQ ID NO:90)	2	+++	+++	+++	+	-

miR-24	UGGCUCAGUUCAGCAGGAACAG (SEQ ID NO:91)	4	++	+++	++	-	-
miR-25	CAUUGCACUUGUCUCGGUCUGA (SEQ ID NO:92)	3	+++	+	++	-	-
miR-26a*	UUCAAGUAAUCCAGGAUAGGCU (SEQ ID NO:93)	2	+	++	+++	-	-
miR-26b*	UUCAAGUAAUUCAGGAUAGGUU (SEQ ID NO:94)	1					-
miR-27	UUCACAGUGGCUAAGUUCCGCU (SEQ ID NO:95)	2	+++	+++	++	-	-
miR-28	AAGGAGCUCACAGUCUAUUGAG (SEQ ID NO:96)	2	+++	+++	-	-	-
miR-29	CUAGCACCAUCUGAAAUCGGUU (SEQ ID NO:97)	2	+	+++	+/-	-	-
miR-30	CUUUCAGUCGGAUGUUUGCAGC (SEQ ID NO:98)	2	+++	+++	+++	-	-
miR-31	GGCAAGAUGCUGGCAUAGCUG (SEQ ID NO:99)	2	+++	-	-	-	-
miR-32	UAUUGCACAUUACUAAGUUGC (SEQ ID NO:100)	1	-	-	-	-	-
miR-33	GUGCAUUGUAGUUGCAUUG (SEQ ID NO:101)	1	-	-	-	-	-
miR-1	UGGAAUGUAAAGAAGUAUGGAG (SEQ ID NO:102)	0	-	-	+	-	-
miR-7	UGGAAGACUAGUGAUUUUGUUGU (SEQ ID NO:103)	0	+	-	+/-	-	+/-
miR-9	UCUUUGGUUAUCUAGCUGUAUGA (SEQ ID NO:104)	0	-	-	-	-	-
miR-10	ACCCUGUAGAUCCGAAUUUGU (SEQ ID NO:105)	0	-	+	-	-	-

= (SEQ ID NO:75)

*Similar miRNA sequences are difficult to distinguish by Northern blotting because of potential cross-hybridization of probes.

Table 3

Mouse miRNAs. The sequences indicated represent the longest miRNA sequences identified by cloning. The 3´-terminus of miRNAs is often truncated by one or two nucleotides. miRNAs that are more than 85% identical in sequence (i.e. share 18 out of 21 nucleotides) or contain 1- or 2-nucleotide internal deletions are referred to by the same gene number followed by a lowercase letter. Minor sequence variations between related miRNAs are generally found near the ends of the miRNA sequence and are thought to not compromise target RNA recognition. Minor sequence variations may also represent A to G and C to U changes, which are accommodated as G-U wobble base pairs during target recognition. miRNAs with the suffix -s or -as indicate RNAs derived from either the 5´-half or the 3´-half of a miRNA precursor. Mouse brains were dissected into midbrain, mb, cortex, cx, cerebellum, cb. The tissues analyzed were heart, ht; liver, lv; small intestine, si; colon, co; cortex, ct; cerebellum, cb; midbrain, mb.

miRNA	sequence (5´ to 3´)	Number of clones							
		ht	lv	sp	si	co	cx	cb	mb
let-7a	UGAGGUAGUAGGUUGUAUAGUU (SEQ ID NO:106)		3			1	1		7
let-7b	UGAGGUAGUAGGUUGUGUGGUU (SEQ ID NO:107)		1	1				2	5
let-7c	UGAGGUAGUAGGUUGUAUGGUU (SEQ ID NO:108)		2				2	5	19
let-7d	AGAGGUAGUAGGUUGCAUAGU (SEQ ID NO:109)	2				2	2		2
let-7e	UGAGGUAGGAGGUUGUAUAGU (SEQ ID NO:110)			1					2
let-7f	UGAGGUAGUAGAUUGUAUAGUU (SEQ ID NO:111)			2				3	3
let-7g	UGAGGUAGUAGUUUGUACAGUA (SEQ ID NO:112)						1	1	2
let-7h	UGAGGUAGUAGUGUGUACAGUU (SEQ ID NO:113)						1	1	

let-7i	UGAGGUAGUAGUUUGUGCU (SEQ ID NO:114)						1	1	
miR-1b	UGGAAUGUAAAGAAGUAUGUAA (SEQ ID NO:115)	4	2						1
miR-1c	UGGAAUGUAAAGAAGUAUGUAC (SEQ ID NO:116)	7							
miR-1d	UGGAAUGUAAAGAAGUAUGUAUU (SEQ ID NO:117)	16							1
miR-9	UCUUUGGUUAUCUAGCUGUAUGA (SEQ ID NO:118)						3	4	4
miR-15a	UAGCAGCACAUAAUGGUUUGUG (SEQ ID NO:119)	1							2
miR-15b	UAGCAGCACAUCAUGGUUUACA (SEQ ID NO:120)	1							
miR-16	UAGCAGCACGUAAAUAUUGGCG (SEQ ID NO:121)	1			1	2	1	2	3
miR-18	UAAGGUGCAUCUAGUGCAGAUA (SEQ ID NO:122)			1					
miR-19b	UGUGCAAAUCCAUGCAAAACUGA (SEQ ID NO:123)			1					
miR-20	UAAAGUGCUUAUAGUGCAGGUAG (SEQ ID NO:124)					1			
miR-21	UAGCUUAUCAGACUGAUGUUGA (SEQ ID NO:125)	1		1	2	1			
miR-22	AAGCUGCCAGUUGAAGAACUGU (SEQ ID NO:126)	2	1		1			1	2
miR-23a	AUCACAUUGCCAGGGAUUUCC (SEQ ID NO:127)	1							
miR-23b	AUCACAUUGCCAGGGAUUACCAC (SEQ ID NO:128)						1		
miR-24	UGGCUCAGUUCAGCAGGAACAG (SEQ ID NO:129)	1				1	1		1
miR-26a	UUCAAGUAAUCCAGGAUAGGCU (SEQ ID NO:130)							3	2
miR-26b	UUCAAGUAAUUCAGGAUAGGUU (SEQ ID NO:131)		2				4	1	
miR-27a	UUCACAGUGGCUAAGUUCCGCU (SEQ ID NO:132)	1		2		1	1	2	1
miR-27b	UUCACAGUGGCUAAGUUCUG (SEQ ID NO:133)								1
miR-29a	CUAGCACCAUCUGAAAUCGGUU (SEQ ID NO:134)	1				1		1	
miR-29b/miR-102	UAGCACCAUUUGAAAUCAGUGUU (SEQ ID NO:135)	1				1	5		3
miR-29c/	UAGCACCAUUUGAAAUCGGUUA (SEQ ID NO:136)	1					3		1

miR-30a-s/miR-97	UGUAAACAUCCUCGACUGGAAGC (SEQ ID NO:137)			1			1		1
miR-30a-as[a]	CUUUCAGUCGGAUGUUUGCAGC (SEQ ID NO:138)							1	
miR-30b	UGUAAACAUCCUACACUCAGC (SEQ ID NO:139)			1				2	
miR-30c	UGUAAACAUCCUACACUCUCAGC (SEQ ID NO:140)	2					1	1	
miR-30d	UGUAAACAUCCCCGACUGGAAG (SEQ ID NO:141)		1						
miR-99a/miR-99	ACCCGUAGAUCCGAUCUUGU (SEQ ID NO:142)						1		
miR-99b	CACCCGUAGAACCGACCUUGCG (SEQ ID NO:143)							1	
miR-101	UACAGUACUGUGAUAACUGA (SEQ ID NO:144)						2	1	1
miR-122a	UGGAGUGUGACAAUGGUGUUUGU (SEQ ID NO:145)		3						
miR-122b	UGGAGUGUGACAAUGGUGUUUGA (SEQ ID NO:146)		11						
miR-122a,b	UGGAGUGUGACAAUGGUGUUUG (SEQ ID NO:147)		23						
miR-123	CAUUAUUACUUUUGGUACGCG (SEQ ID NO:148)	1	2						
miR-124a[b]	UUAAGGCACGCGG-UGAAUGCCA (SEQ ID NO:149)				1		37	41	24
miR-124b	UUAAGGCACGCGGGUGAAUGC (SEQ ID NO:150)						1	3	
miR-125a	UCCCUGAGACCCUUUAACCUGUG (SEQ ID NO:151)						1	1	
miR-125b	UCCCUGAGACCCU--AACUUGUGA (SEQ ID NO:152)						1		
miR-126	UCGUACCGUGAGUAAUAAUGC (SEQ ID NO:153)	4						1	
miR-127	UCGGAUCCGUCUGAGCUUGGCU (SEQ ID NO:154)							1	
miR-128	UCACAGUGAACCGGUCUCUUUU (SEQ ID NO:155)						2	2	2
miR-129	CUUUUUCGGUCUGGGCUUGC (SEQ ID NO:156)							1	
miR-130	CAGUGCAAUGUUAAAAGGGC (SEQ ID NO:157)							1	
miR-131	UAAAGCUAGAUAACCGAAAGU (SEQ ID NO:158)						1	1	1
miR-132	UAACAGUCUACAGCCAUGGUCGU (SEQ ID NO:159)							1	

miR-133	UUGGUCCCCUUCAACCAGCUGU (SEQ ID NO:160)	4					1		
miR-134	UGUGACUGGUUGACCAGAGGGA (SEQ ID NO:161)						1		
miR-135	UAUGGCUUUUUAUUCCUAUGUGAA (SEQ ID NO:162)						1		
miR-136	ACUCCAUUUGUUUUGAUGAUGGA (SEQ ID NO:163)						1		
miR-137	UAUUGCUUAAGAAUACGCGUAG (SEQ ID NO:164)						1		1
miR-138	AGCUGGUGUUGUGAAUC (SEQ ID NO:165)						1		
miR-139	UCUACAGUGCACGUGUCU (SEQ ID NO:166)					1	1		
miR-140	AGUGGUUUUACCCUAUGGUAG (SEQ ID NO:167)				1				
miR-141	AACACUGUCUGGUAAAGAUGG (SEQ ID NO:168)			1	1		1		
miR-142-s	CAUAAAGUAGAAAGCACUAC (SEQ ID NO:169)				1	1			
miR-142-as[b]	UGUAGUGUUUCCUACUUUAUGG (SEQ ID NO:170)			1	1	6			
miR-143	UGAGAUGAAGCACUGUAGCUCA (SEQ ID NO:171)	3		7			2		1
miR-144	UACAGUAUAGAUGAUGUACUAG (SEQ ID NO:172)	2				1			
miR-145	GUCCAGUUUUCCCAGGAAUCCCUU (SEQ ID NO:173)	1							
miR-146	UGAGAACUGAAUUCCAUGGGUUU (SEQ ID NO:174)	1							
miR-147	GUGUGUGGAAAUGCUUCUGCC (SEQ ID NO:175)			1					
miR-148	UCAGUGCACUACAGAACUUUGU (SEQ ID NO:176)			1					
miR-149	UCUGGCUCCGUGUCUUCACUCC (SEQ ID NO:177)	1							
miR-150	UCUCCCAACCCUUGUACCAGUGU (SEQ ID NO:178)					1			
miR-151	CUAGACUGAGGCUCCUUGAGGU (SEQ ID NO:179)					1			
miR-152	UCAGUGCAUGACAGAACUUGG (SEQ ID NO:180)					1			
miR-153	UUGCAUAGUCACAAAAGUGA (SEQ ID NO:181)								1
miR-154	UAGGUUAUCCGUGUUGCCUUCG (SEQ ID NO.182)								1

miR-155	UUAAUGCUAAUUGUGAUAGGGG (SEQ ID NO:183)	1

[a]The originally described miR-30 was renamed to miR-30a-as in order to distinguish it from the miRNA derived from the opposite strand of the precursor encoded by the mir-30a gene. miR-30a-s is equivalent to miR-97 [46].
[b]A 1-nt length heterogeneity is found on both 5′ and 3′ end. The 22-nt miR sequence is shown, but only 21-nt miRNAs were cloned.

Table 4

Mouse and human miRNAs. The sequences indicated represent the longest miRNA sequences identified by cloning. The 3′ terminus of miRNAs is often truncated by one or two nucleotides. miRNAs that are more than 85% identical in sequence (i.e. share 18 out of 21 nucleotides) or contain 1- or 2-nucleotide internal deletions are referred to by the same gene number followed by a lowercase letter. Minor sequence variations between related miRNAs are generally found near the ends of the miRNA sequence and are thought to not compromise target RNA recognition. Minor sequence variations may also represent A to G and C to U changes; which are accommodated as G-U wobble base pairs during target recognition. Mouse brains were dissected into midbrain, mb, cortex, cx, cerebellum, cb. The tissues analyzed were lung, ln; liver, lv; spleen, sp; kidney, kd; skin, sk; testis, ts; ovary, ov; thymus, thy; eye, ey; cortex, ct; cerebellum, cb; midbrain, mb. The human osteosarcoma cells SAOS-2 cells contained an inducible p53 gene (p53-, uninduced p53; p53+, induced p53); the differences in miRNAs identified from induced and uninduced SAOS cells were not statistically significant.

miRNA	Sequence (5′ to 3′)	number of clones											
		mouse tissues									human SAOS-2 cells		
		ln	lv	sp	kd	sk	ts	ov	thy	ey	p53-	p53+	
miR-C1	AACAUUCAACGCUGUCGGUGAGU	1				1				2			(SEQ ID NO.184)
miR-C2	UUUGGCAAUGGUAGAACUCACA									1			(SEQ ID NO.185)
miR-C3	UAUGGCACUGGUAGAAUUCACUG									1			(SEQ ID NO.186)
miR-C4	CUUUUUGCGGUCUGGGCUUGUU						1			1	1		(SEQ ID NO.187)
miR-C5	UGGACGGAGAACUGAUAAGGGU									2			(SEQ ID NO.188)
miR-C6	UGGAGAGAAAGGCAGUUC									1			(SEQ ID NO.189)
miR-C7	CAAAGAAUUCUCCUUUUGGGCUU									1	1		(SEQ ID NO.190)
miR-C8	UCGUGUCUUGUGUUGCAGCCGG				1								(SEQ ID NO.191)
miR-C9	UAACACUGUCUGGUAACGAUG				1								(SEQ ID NO.192)
miR-C10	CAUCCCUUGCAUGGUGGAGGGU				1								(SEQ ID NO.193)
miR-C11	GUGCCUACUGAGCUGACAUCAGU				1								(SEQ ID NO.194)
miR-C12	UGAUAUGUUUGAUAUAUUAGGU				2								(SEQ ID NO.195)
miR-C13	CAACGGAAUCCCAAAAGCAGCU			2	1								(SEQ ID NO.196)
miR-C14	CUGACCUAUGAAUUGACA		2		1								(SEQ ID NO.197)

miR-C15	UACCACAGGGUAGAACCACGGA		1								(SEQ ID NO.198)
miR-C16	AACUGGCCUACAAAGUCCCAG		1								(SEQ ID NO.199)
miR-C17	UGUAACAGCAACUCCAUGUGGA		1								(SEQ ID NO.200)
miR-C18	UAGCAGCACAGAAAUAUUGGC	2	1	1							(SEQ ID NO.201)
miR-C19	UAGGUAGUUUCAUGUUGUUGG								1		(SEQ ID NO.202)
miR-C20	UUCACCACCUUCUCCACCCAGC								1	1	(SEQ ID NO.203)
miR-C21	GGUCCAGAGGGGAGAUAGG									1	(SEQ ID NO.204)
miR-C22	CCCAGUGUUCAGACUACCUGUU									1	(SEQ ID NO.205)
miR-C23	UAAUACUGCCUGGUAAUGAUGAC	2			1						(SEQ ID NO.206)
miR-C24	UACUCAGUAAGGCAUUGUUCU				1						(SEQ ID NO.207)
miR-C25	AGAGGUAUAGCGCAUGGGAAGA				1						(SEQ ID NO.208)
miR-C26	UGAAAUGUUUAGGACCACUAG				1						(SEQ ID NO.209)
miR-C27	UUCCCUUUGUCAUCCUAUGCCUG							1			(SEQ ID NO.210)
miR-C28	UCCUUCAUUCCACCGGAGUCUG			1							(SEQ ID NO.211)
miR-C29	GUGAAAUGUUUAGGACCACUAGA			2							(SEQ ID NO.212)
miR-C30	UGGAAUGUAAGGAAGUGUGUGG			2							(SEQ ID NO.213)
miR-C31	UACAGUAGUCUGCACAUUGGUU			1							(SEQ ID NO.214)
miR-C32	CCCUGUAGAACCGAAUUUGUGU			1		1					(SEQ ID NO.215)
miR-C33	AACCCGUAGAUCCGAACUUGUGAA			1							(SEQ ID NO.216)
miR-C34	GCUUCUCCUGGCUCUCCUCCCUC						1				(SEQ ID NO.217)

Table 5

D. melanogaster miRNA sequences and genomic location. The sequences given represent the most abundant, and typically longest miRNA sequences identified by cloning. It was frequently observed that miRNAs vary in length by one or two nucleotides at their 3′-terminus. From 222 short RNAs sequenced; 69 (31%) corresponded to miRNAs, 103 (46%) to already characterized functional RNAs (rRNA, 7SL RNA, tRNAs), 30 (14%) to transposon RNA fragments, and 20 (10%) sequences with no database entry. RNA sequences with a 5′-guanosine are likely to be underrepresented due to the cloning procedure (8). miRNA homologs found in other species are indicated. Chromosomal location (chr.) and GenBank accession numbers (acc. nb.) are indicated. No ESTs matching miR-1 to miR-14 were detectable by database searching.

miRNA	sequence (5′ to 3′)	chr., acc. nb.	remarks
miR-1	UGGAAUGUAAAGAAGUAUGGAG (SEQ ID NO:58)	2L, AE003667	homologs: *C. briggsae*, G20U, AC87074; *C.elegans* G20U, U97405; mouse, G20U, G22U, AC020867; human, chr. 20, G20U, G22U, AL449263; ESTs: zebrafish, G20U, G22U, BF157-601; cow, G20U, G22U, BE722-224; human, G20U, G22U, AI220268
miR-2a	UAUCACAGCCAGCUUUGAUGAGC (SEQ ID NO:59)	2L, AE003663	2 precursor variants clustered with a copy of *mir-2b*
miR-2b	UAUCACAGCCAGCUUUGAGGAGC (SEQ ID NO:60)	2L, AE003620 2L, AE003663	2 precursor variants
miR-3	UCACUGGGCAAAGUGUGUCUCA (SEQ ID NO:61)	2R, AE003795	in cluster *mir-3* to *mir-6*
miR-4	AUAAAGCUAGACAACCAUUGA (SEQ ID NO:62)	2R, AE003795	in cluster *mir-3* to *mir-6*

miR-5	AAAGGAACGAUCGUUGUGAUAUG (SEQ ID NO:63)	2R, AE003795	in cluster *mir-3* to *mir-6*
miR-6	UAUCACAGUGGCUGUUCUUUUU (SEQ ID NO:64)	2R, AE003795	in cluster *mir-3* to *mir-6* with 3 variants
miR-7	UGGAAGACUAGUGAUUUUGUUGU (SEQ ID NO:65)	2R, AE003791	homologs: human, chr. 19 AC006537, EST BF373391; mouse chr. 17 AC026385, EST AA881786
miR-8	UAAUACUGUCAGGUAAAGAUGUC (SEQ ID NO:66)	2R, AE003805	
miR-9	UCUUUGGUUAUCUAGCUGUAUGA (SEQ ID NO:67)	3L, AE003516	homologs: mouse, chr. 19, AF155142; human, chr. 5, AC026701, chr. 15, AC005316
miR-10	ACCCUGUAGAUCCGAAUUUGU (SEQ ID NO:68)	AE001574	homologs: mouse, chr 11, AC011194; human, chr. 17, AF287967
miR-11	CAUCACAGUCUGAGUUCUUGC (SEQ ID NO:69)	3R, AE003735	intronic location
miR-12	UGAGUAUUACAUCAGGUACUGGU (SEQ ID NO:70)	X, AE003499	intronic location
miR-13a	UAUCACAGCCAUUUUGACGAGU (SEQ ID NO:71)	3R, AE003708 X, AE003446	*mir-13a* clustered with *mir-13b* on chr. 3R
miR-13b	UAUCACAGCCAUUUUGAUGAGU (SEQ ID NO:72)	3R, AE003708	*mir-13a* clustered with *mir-13b* on chr. 3R
miR-14	UCAGUCUUUUUCUCUCUCCUA (SEQ ID NO:73)	2R, AE003833	no signal by Northern analysis

Table 6

Human miRNA sequences and genomic location. From 220 short RNAs sequenced, 100 (45%) corresponded to miRNAs, 53 (24%) to already characterized functional RNAs (rRNA, snRNAs, tRNAs), and 67 (30%) sequences with no database entry. For legend, see Table 1.

miRNA	sequence (5′ to 3′)	chr. or EST, acc. nb.	remarks*
let-7a	UGAGGUAGUAGGUUGUAUAGUU (SEQ ID NO:75)	9, AC007924, 11, AP001359, 17, AC087784, 22, AL049853	sequences of chr 9 and 17 identical and clustered with *let-7f*, homologs: *C. elegans*, AF274345; *C. briggsae*, AF210771, *D. melanogaster*, AE003659
let-7b	UGAGGUAGUAGGUUGUGUGGUU (SEQ ID NO:76)	22, AL049853†, ESTs, AI382133, AW028822	homologs: mouse, EST AI481799; rat, EST, BE120662
let-7c	UGAGGUAGUAGGUUGUAUGGUU (SEQ ID NO:77)	21, AP001667	Homologs: mouse, EST, AA575575
let-7d	AGAGGUAGUAGGUUGCAUAGU (SEQ ID NO:78)	17, AC087784, 9, AC007924	identical precursor sequences
let-7e	UGAGGUAGGAGGUUGUAUAGU (SEQ ID NO:79)	19, AC018755	
let-7f	UGAGGUAGUAGAUUGUAUAGUU (SEQ ID NO:80)	9, AC007924, 17, AC087784, X, AL592046	sequences of chr 9 and 17 identical and clustered with *let-7a*
miR-15	UAGCAGCACAUAAUGGUUUGUG (SEQ ID NO:81)	13, AC069475	in cluster with *mir-16* homolog
miR-16	UAGCAGCACGUAAAUAUUGGCG (SEQ ID NO:82)	13, AC069475	in cluster with *mir-15* homolog

miR-17	ACUGCAGUGAAGGCACUUGU (SEQ ID NO:83)	13, AL138714	in cluster with *mir-17* to *mir-20*
miR-18	UAAGGUGCAUCUAGUGCAGAUA (SEQ ID NO:84)	13, AL138714	in cluster with *mir-17* to *mir-20*
miR-19a	UGUGCAAAUCUAUGCAAAACUGA (SEQ ID NO:85)	13, AL138714	in cluster with *mir-17* to *mir-20*
miR-19b	UGUGCAAAUCCAUGCAAAACUGA (SEQ ID NO:86)	13, AL138714, X, AC002407	in cluster with *mir-17* to *mir-20*
miR-20	UAAAGUGCUUAUAGUGCAGGUA (SEQ ID NO:87)	13, AL138714	in cluster with *mir-17* to *mir-20*
miR-21	UAGCUUAUCAGACUGAUGUUGA (SEQ ID NO:88)	17, AC004686, EST, BF326048	homologs: mouse, EST, AA209594
miR-22	AAGCUGCCAGUUGAAGAACUGU (SEQ ID NO:89)	ESTs, AW961681†, AA456477, AI752503, BF030303, HS1242049	human ESTs highly similar; homologs: mouse, ESTs, e.g. AA823029; rat, ESTs, e.g. BF543690
miR-23	AUCACAUUGCCAGGGAUUUCC (SEQ ID NO:90)	19, AC020916	homologs: mouse, EST, AW124037;rat, EST, BF402515
miR-24	UGGCUCAGUUCAGCAGGAACAG (SEQ ID NO:91)	9, AF043896, 19, AC020916	homologs: mouse, ESTs, AA111466, AI286629; pig, EST, BE030976
miR-25	CAUUGCACUUGUCUCGGUCUGA (SEQ ID NO:92)	7, AC073842, EST, BE077684	human chr 7 and EST identical; highly similar precursors in mouse ESTs (e.g. AI595464); fish precursor different STS: G46757
miR-26a	UUCAAGUAAUCCAGGAUAGGCU (SEQ ID NO:93)	3, AP000497	

miR-26b	UUCAAGUAAUUCAGGAUAGGUU (SEQ ID NO:94)	2, AC021016	
miR-27	UUCACAGUGGCUAAGUUCCGCU (SEQ ID NO:95)	19, AC20916	U22C mutation in human genomic sequence
miR-28	AAGGAGCUCACAGUCUAUUGAG (SEQ ID NO:96)	3, AC063932	
miR-29	CUAGCACCAUCUGAAAUCGGUU (SEQ ID NO:97)	7, AF017104	
miR-30	CUUUCAGUCGGAUGUUUGCAGC (SEQ ID NO:98)	6, AL035467	
miR-31	GGCAAGAUGCUGGCAUAGCUG (SEQ ID NO:99)	9, AL353732	
miR-32	UAUUGCACAUUACUAAGUUGC (SEQ ID NO:100)	9, AL354797	not detected by Northern blotting
miR-33	GUGCAUUGUAGUUGCAUUG (SEQ ID NO:101)	22, Z99716	not detected by Northern blotting

*If several ESTs were retrieved for one organism in the database, only those with different precursor sequences are listed.

†precursor structure shown in Fig. 4.

Claims

1. Isolated nucleic acid molecule comprising

   (a) a nucleotide sequence as shown in Table 1, Table 2, Table 3 or Table 4 or a precursor thereof as shown in Figure 3, Figure 4 or Figure 7.

   (b) a nucleotide sequence which is the complement of (a),

   (c) a nucleotide sequence which has an identity of at least 80% to a sequence of (a) or (b) and/or

   (d) a nucleotide sequence which hybridizes under stringent conditions to a sequence of (a), (b) and/or (c).

2. The nucleic acid molecule of claim 1, wherein the identity of sequence (c) is at least 90%.

3. The nucleic acid molecule of claim 1, wherein the identity of sequence (c) is at least 95%.

4. The nucleic acid molecule of any one of claims 1-3, which is selected from miR 1-14 as shown in Table 1 or miR 15-33 as shown in Table 2 or miR 1-155 as shown in Table 3 or miR-C1-34 as shown in Table 4 or a complement thereof.

5. The nucleic acid molecule of any one of claims 1-3, which is selected from mir 1-14 as shown in Figure 3 or let 7a-7f or mir 15-33, as shown in Figure 4 or let 7a-i or mir 1-155 or mir-c1-34, as shown in Figure 7 or a complement thereof.

6. The nucleic acid molecule of any one of claims 1-4 which is a miRNA molecule or an analog thereof having a length of from 18-25 nucleotides.

7. The nucleic acid molecule of any one of claims 1-3 or 5, which is a miRNA precursor molecule having a length of 60-80 nucleotides or a DNA molecule coding therefor.

8. The nucleic acid molecule of any one of claims 1-7, which is single-stranded.

9. The nucleic acid molecule of any one of claims 1-7, which is at least partially double-stranded.

10. The nucleic acid molecule of any one of claims 1-9, which is selected from RNA, DNA or nucleic acid analog molecules.

11. The nucleic acid molecule of claim 10, which is a molecule containing at least one modified nucleotide analog.

12. The nucleic molecule of claim 10 which is a recombinant expression vector.

13. A pharmaceutical composition containinig as an active agent at least one nucleic acid molecule of any one of claims 1-12 and optionally a pharmaceutically acceptable carrier.

14. The composition of claim 13 for diagnostic applications.

15. The composition of claim 13 for therapeutic applications.

16. The composition of any one of claims 13-15 as a marker or a modulator for developmental or pathogenic processes.

17. The composition of claim 13 as a marker or modulator of developmental disorders, particularly cancer, such a B-cell chronic leukemia.

18. The composition of any one of claims 13-15 as a marker or modulator of gene expression.

19. The composition of claim 18 as a marker or modulator of the expression of a gene, which is at least partially complementary to said nucleic acid molecule.

20. A method of identifying microRNA molecules or precursor molecules thereof comprising ligating 5'- and 3'-adapter molecules to the ends of a size-fractionated RNA population, reverse transcribing said adapter-containing RNA population and characterizing the reverse transcription products.

Fig. 1 A

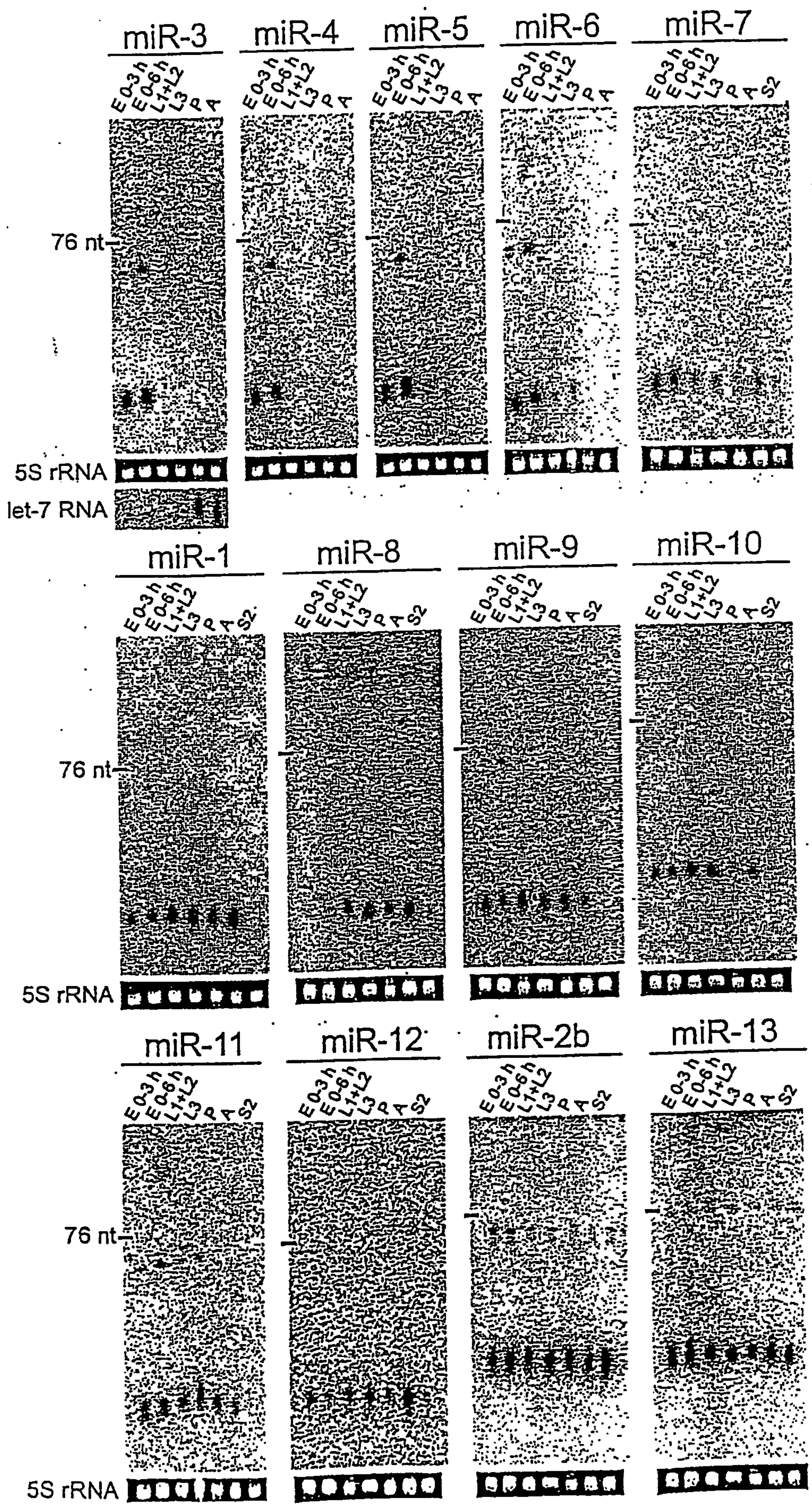
miR-3
miR-4
miR-5
miR-6
miR-7
E 0-3 h
E 0-6 h
L1+L2
L3
P
A
S2
76 nt
5S rRNA
let-7 RNA
miR-1
miR-8
miR-9
miR-10
miR-11
miR-12
miR-2b
miR-13

Fig. 1 B

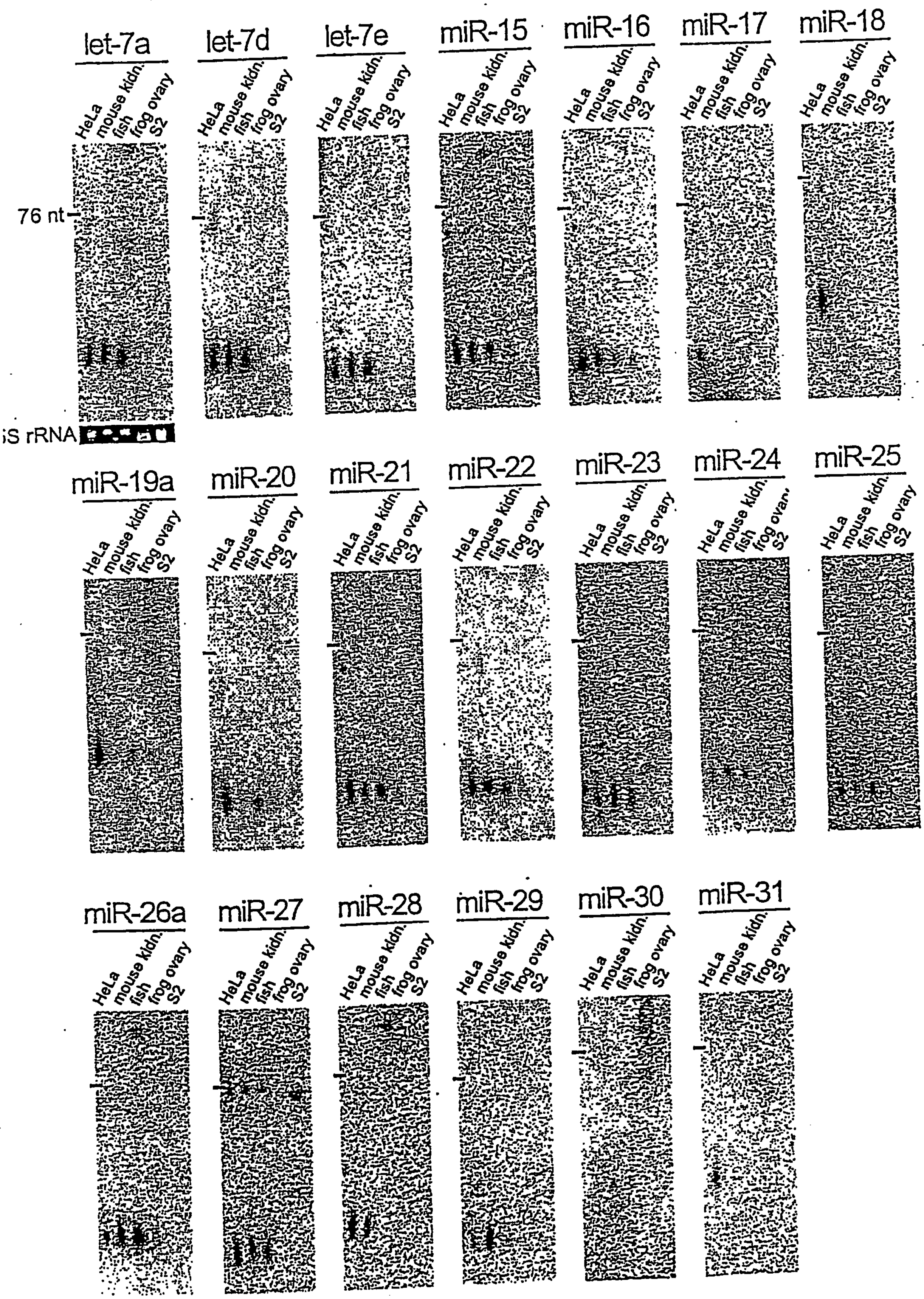
let-7a
let-7d
let-7e
miR-15
miR-16
miR-17
miR-18
HeLa
mouse kidn.
fish
frog ovary
S2
76 nt
5S rRNA
miR-19a
miR-20
miR-21
miR-22
miR-23
miR-24
miR-25
miR-26a
miR-27
miR-28
miR-29
miR-30
miR-31

Fig. 2

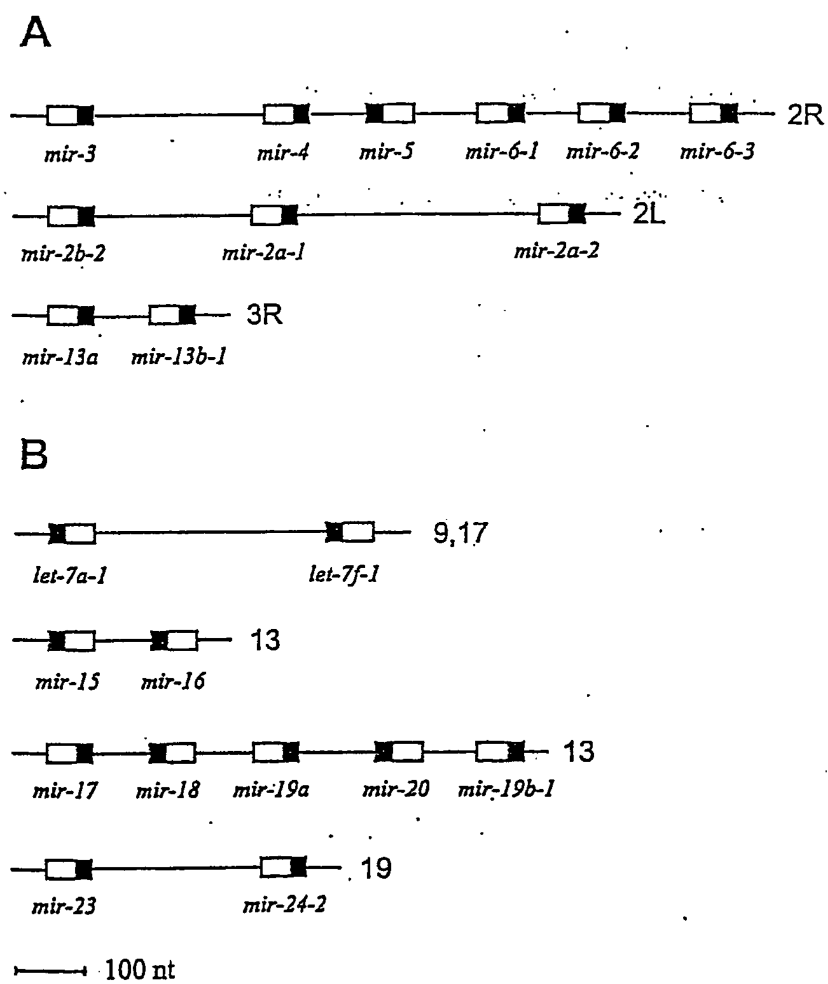
A
mir-3
mir-4
mir-5
mir-6-1
mir-6-2
mir-6-3
2R
mir-2b-2
mir-2a-1
mir-2a-2
2L
mir-13a
mir-13b-1
3R
B
let-7a-1
let-7f-1
9,17
mir-15
mir-16
13
mir-17
mir-18
mir-19a
mir-20
mir-19b-1
13
mir-23
mir-24-2
19
100 nt

Fig. 3

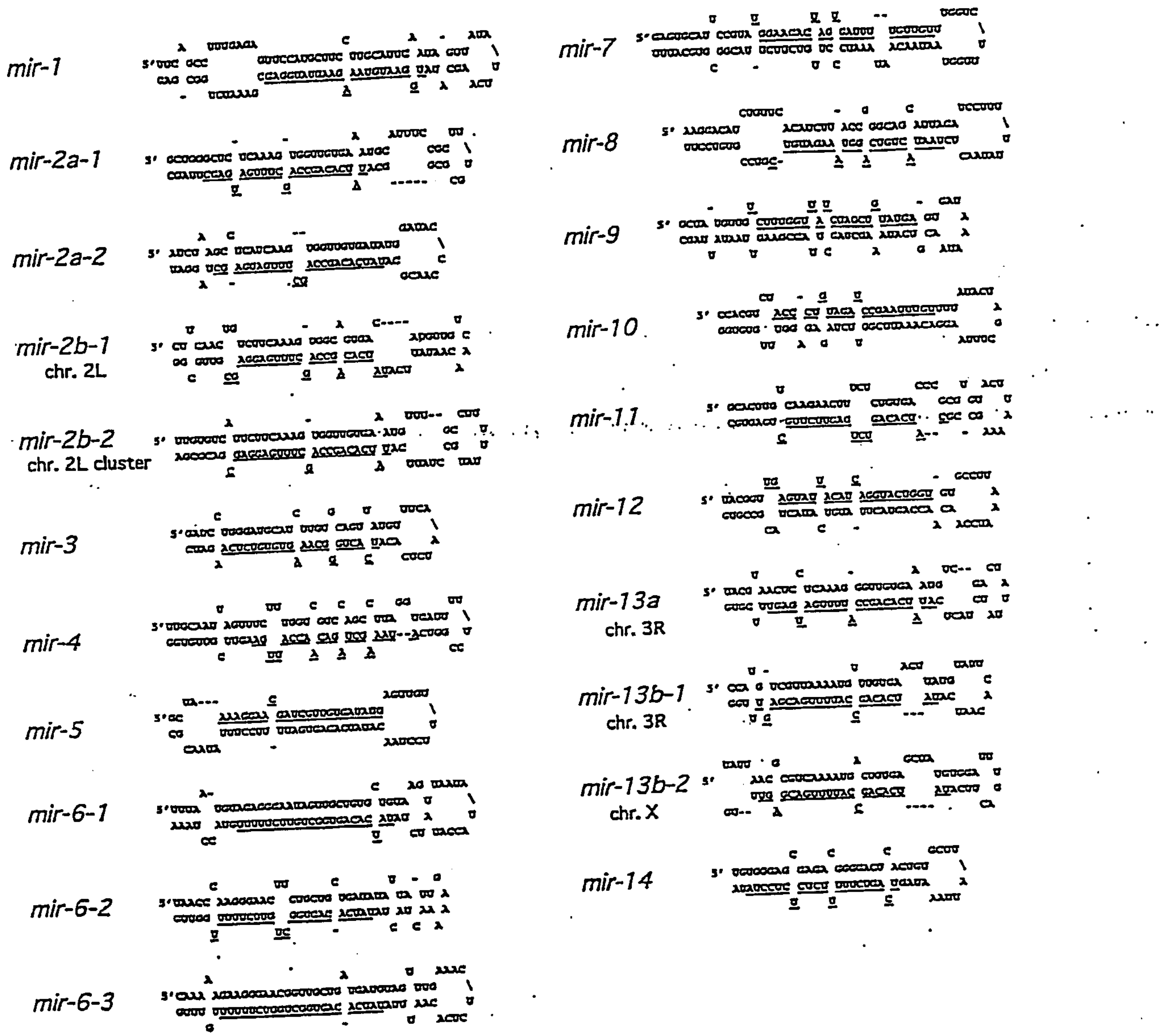
mir-1
mir-2a-1
mir-2a-2
mir-2b-1
chr. 2L
mir-2b-2
chr. 2L cluster
mir-3
mir-4
mir-5
mir-6-1
mir-6-2
mir-6-3
mir-7
mir-8
mir-9
mir-10
mir-11
mir-12
mir-13a
chr. 3R
mir-13b-1
chr. 3R
mir-13b-2
chr. X
mir-14

Fig. 4

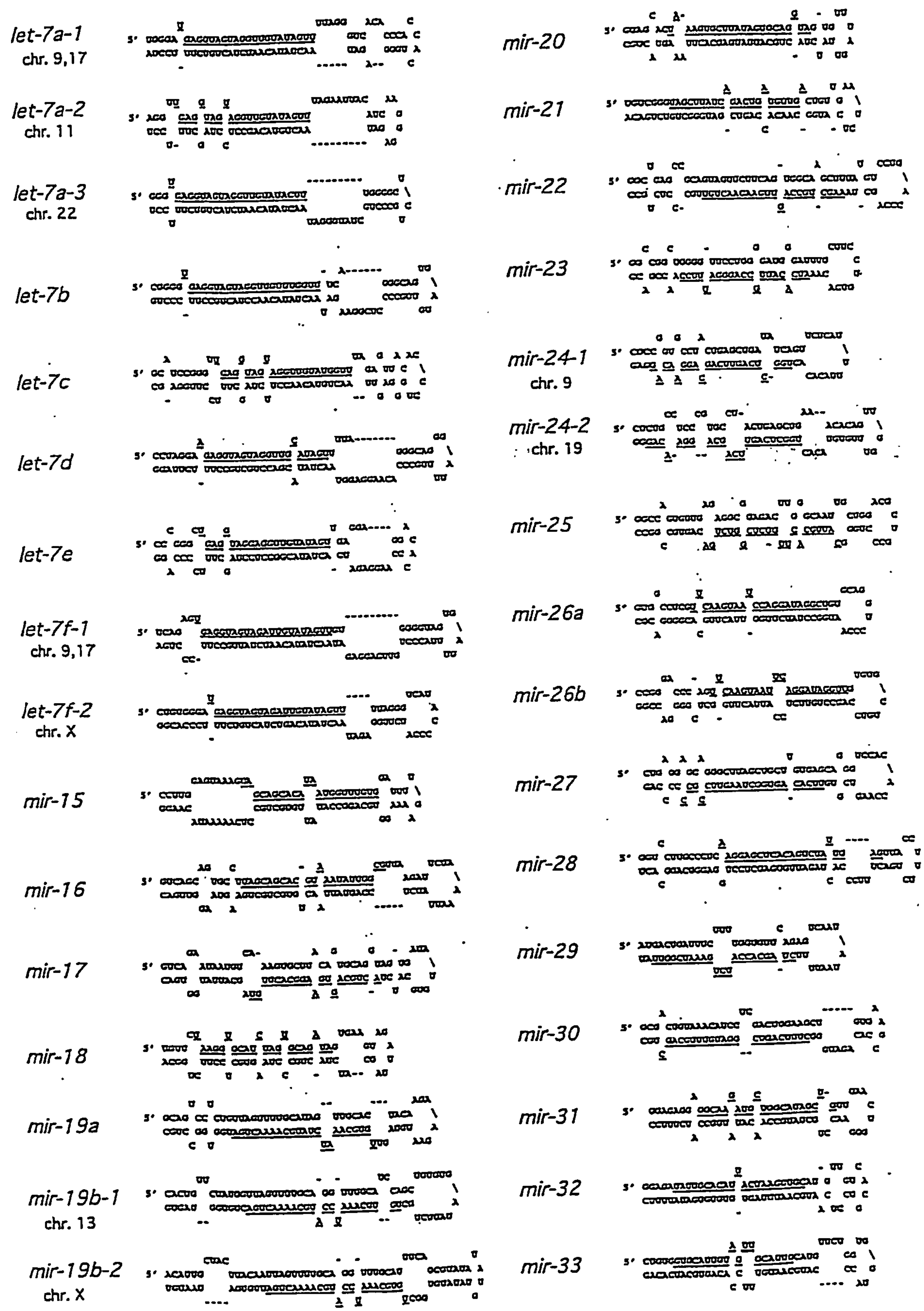
let-7a-1
chr. 9,17
let-7a-2
chr. 11
let-7a-3
chr. 22
let-7b
let-7c
let-7d
let-7e
let-7f-1
chr. 9,17
let-7f-2
chr. X
mir-15
mir-16
mir-17
mir-18
mir-19a
mir-19b-1
chr. 13
mir-19b-2
chr. X
mir-20
mir-21
mir-22
mir-23
mir-24-1
chr. 9
mir-24-2
chr. 19
mir-25
mir-26a
mir-26b
mir-27
mir-28
mir-29
mir-30
mir-31
mir-32
mir-33

Fig. 5

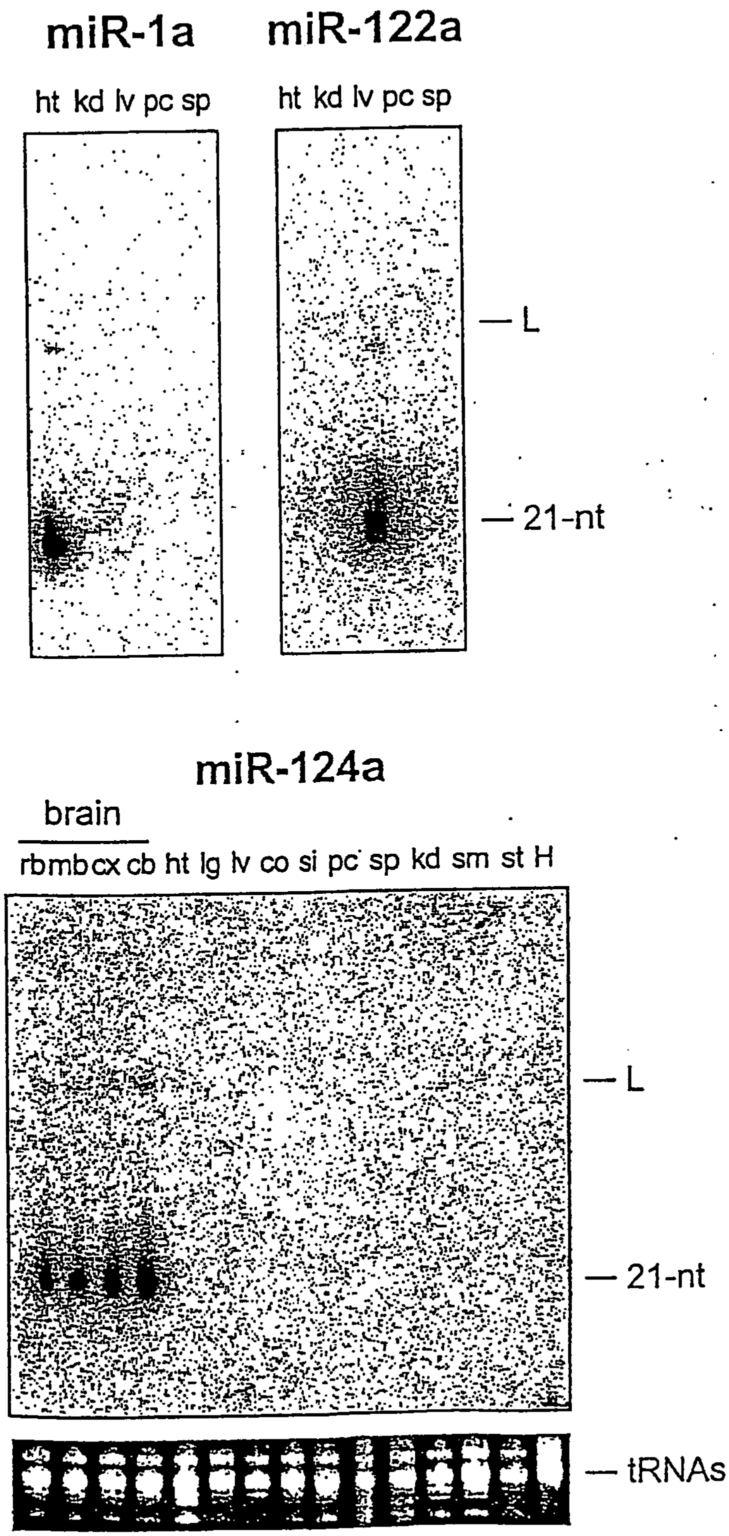
miR-1a
miR-122a
ht kd lv pc sp
ht kd lv pc sp
— L
— 21-nt
miR-124a
brain
rb mb cx cb ht lg lv co si pc sp kd sm st H
— L
— 21-nt
— tRNAs

Fig. 5 (cont.)

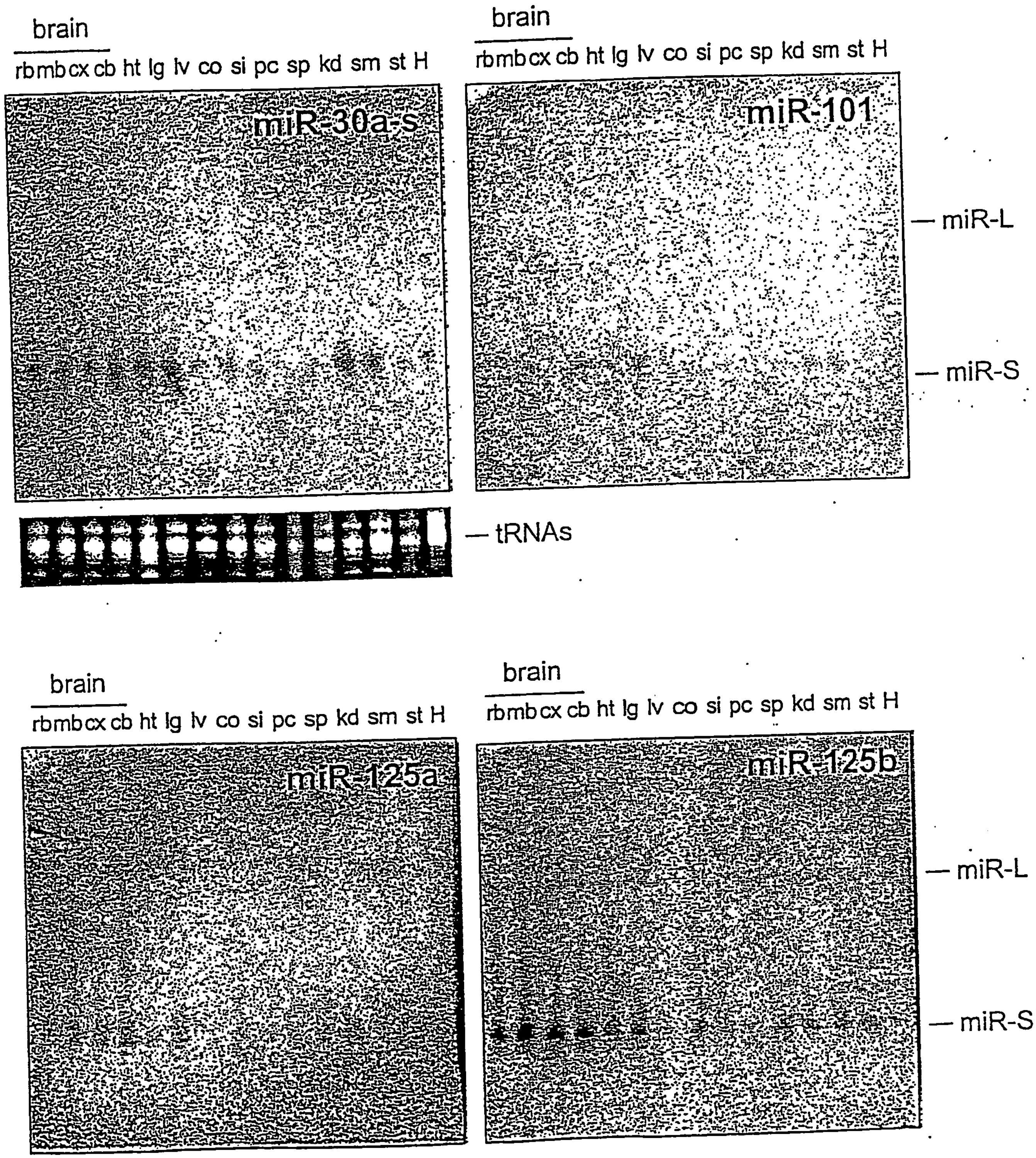
brain
rb mb cx cb ht lg lv co si pc sp kd sm st H
miR-30a-s
tRNAs
brain
rb mb cx cb ht lg lv co si pc sp kd sm st H
miR-101
miR-L
miR-S
brain
rb mb cx cb ht lg lv co si pc sp kd sm st H
miR-125a
brain
rb mb cx cb ht lg lv co si pc sp kd sm st H
miR-125b
miR-L
miR-S

Fig. 5 (cont.)

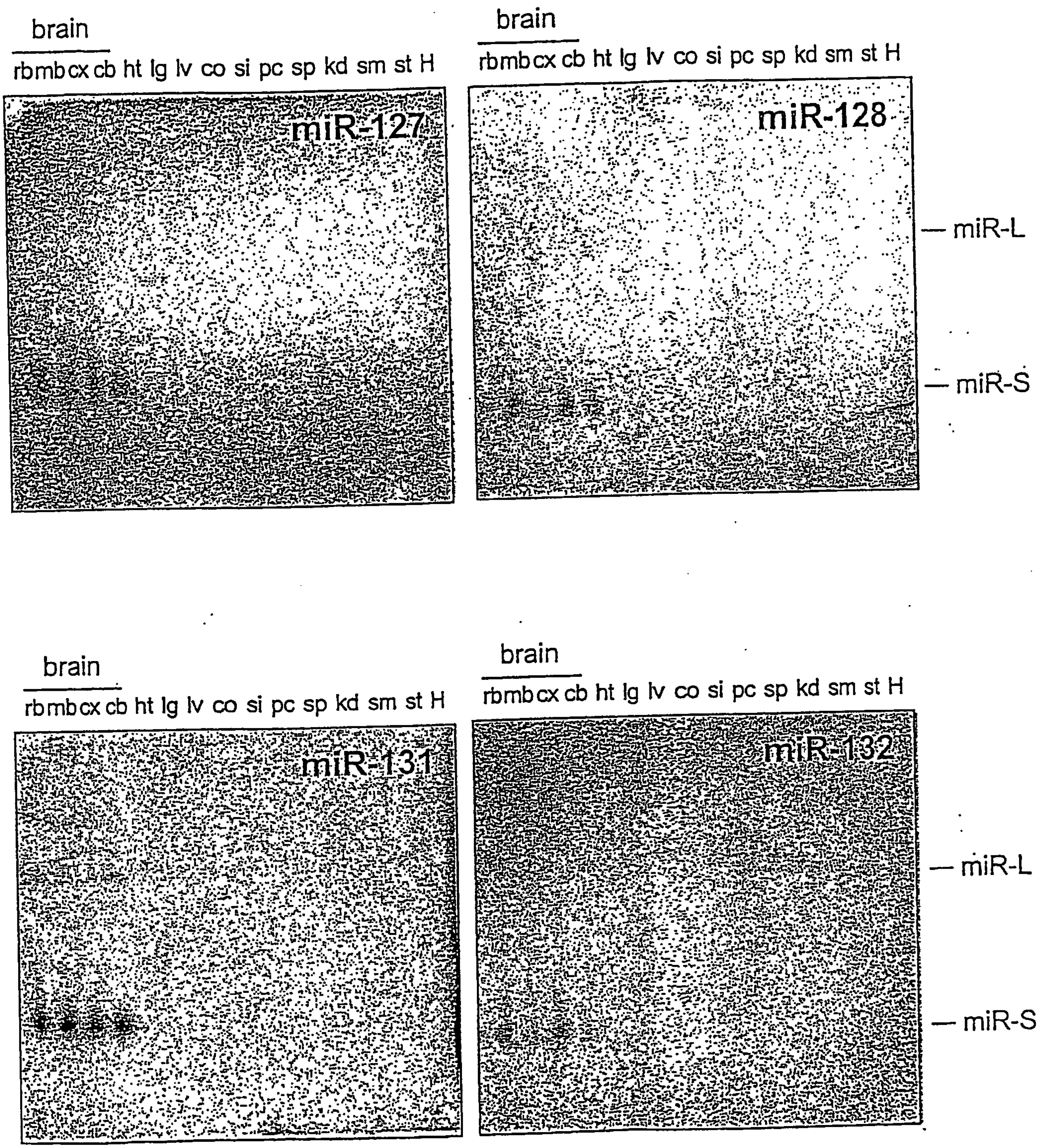
brain
rb mb cx cb ht lg lv co si pc sp kd sm st H
miR-127
brain
rb mb cx cb ht lg lv co si pc sp kd sm st H
miR-128
miR-L
miR-S
brain
rb mb cx cb ht lg lv co si pc sp kd sm st H
miR-131
brain
rb mb cx cb ht lg lv co si pc sp kd sm st H
miR-132
miR-L
miR-S

Fig. 5 (cont.)

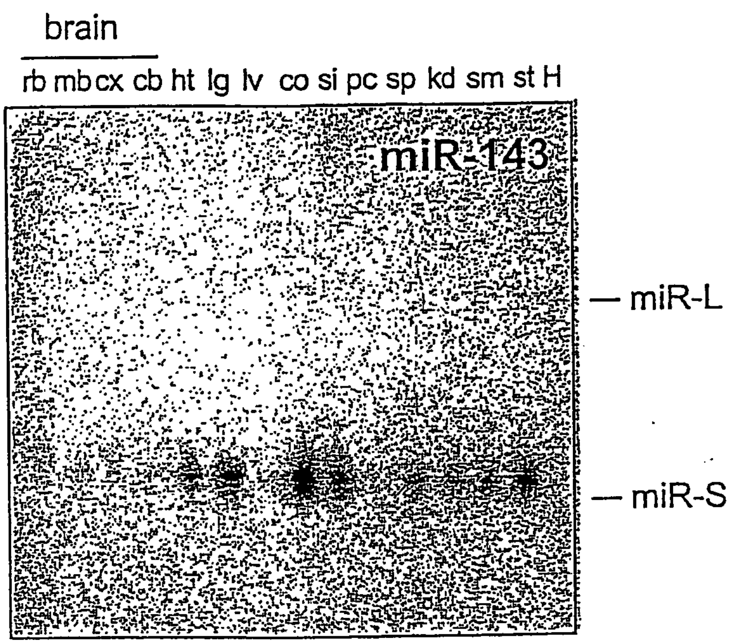

Fig. 6

A

C. elegans lin-4	UCCCUGAGACCUC--AAG-UGUGA
D. melanogaster miR-125	UCCCUGAGACCCU--AACUUGUGA
M. musculus/H. sapiens miR-125b	UCCCUGAGACCCU--AACUUGUGA
M. musculus/H. sapiens miR-125a	UCCCUGAGACCCUUUAACCUGUGA

B

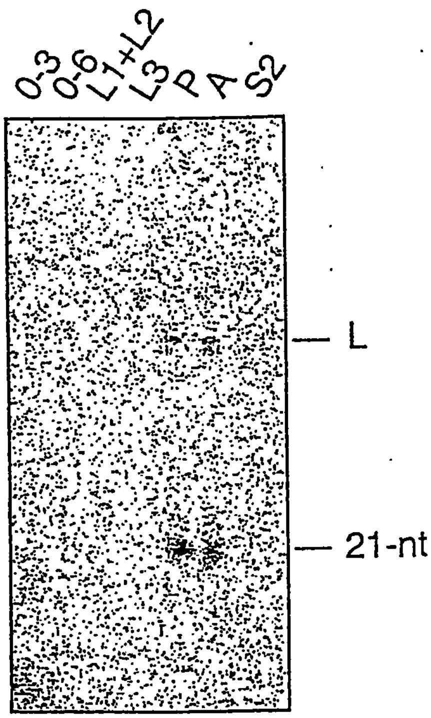

Fig. 7

name	sequence	structure
let-7a-1	UGAGGUAGUAGGUUGUAUAGUU	UG U UUAGG ACA C CAC UGGGA GAGGUAGUAGGUUGUAUAGUU GUC CCCA C GUG AUCCU UUCUGUCAUCUAACAUAUCAA UAG GGGU A CA - ----- A-- C
let-7a-2	UGAGGUAGUAGGUUGUAUAGUU	UU G U UAGAAUUAC AA AGG GAG UAG AGGUUGUAUAGUU AUC G UCC UUC AUC UCCGACAUGUCAA UAG G U- G C ---------- AG
let-7a-3	UGAGGUAGUAGGUUGUAUAGUU	U ---------- U GGG GAGGUAGUAGGUUGUAUAGUU UGGGGC \\ UCC UUCUGUCAUCUAACAUAUCAA GUCCCG C U UAGGGUAUC U
let-7b	UGAGGUAGUAGGUUGUGUGGUU	GG U - A------- UG CGGGG GAGGUAGUAGGUUGUGUGGUU UC GGGCAG \\ GUCCC UUCCGUCAUCCAACAUAUCAA AG CCCGUU A -- - U AAGGCUC GU
let-7c	UGAGGUAGUAGGUUGUAUGGUU	A UU G U UA G UA AC GC UCCGGG GAG UAG AGGUUGUAUGGUU GA U C \\ CG AGGUUC UUC AUC UCCAACAUGUCAA UU A G C - CU G U -- G GG UC
let-7d	AGAGGUAGUAGGUUGCAUAGU	A C UUA-------- GG CCUAGGA GAGGUAGUAGGUUG AUAGUU GGGCAG \\ GGAUUCU UUCCGUCGUCCAGC UAUCAA CCCGUU A - A UGGAGGAACA UU
let-7e	UGAGGUAGGAGGUUGUAUAGU	C CU G U GGA---- A CC GGG GAG UAGGAGGUUGUAUAGU GA GG C GG CCC UUC AUCCUCCGGCAUAUCA CU CC A A CU G - AGAGGAA C

Fig. 7 (cont.)

let-7f-1	UGAGGUAGUAGAUUGUAUAGUU	AGU ---------- UG UCAG GAGGUAGUAGAUUGUAUAGUUGU GGGGUAG \\ AGUC UUCCGUUAUCUAACAUAUCAAUA UCCCAUU A CC- GAGGACUUG UU
let-7f-2	UGAGGUAGUAGAUUGUAUAGUU	U ---- UCAU CUGUGGGA GAGGUAGUAGAUUGUAUAGUU UUAGGG A GGCACCCU UUCUGUCAUCUGACAUAUCAA GGUUCU C - UAGA ACCC
let-7g	UGAGGUAGUAGUUUGUACAGUA	A U A UGAGG A- A A CC GGC GAGGUAGU GUUUGUACAGUU GUCU UG UACC C GG CCG UUCCGUCA CGGACAUGUCAA UAGA AC AUGG C A - C ----- GG - C
let-7h	UGAGGUAGUAGUGUGUACAGUU	
let-7i	UGAGGUAGUAGUUUGUGCU	U U --------- U UGUG CUGGC GAGGUAGUAGUUUGUGC GUU GG CGGGU \\ GAUCG UUCCGUCAUCGAACGCG CAA UC GCCCG A - U UAGAGGUG - UUAC
miR-1	UGGAAUGUAAAGAAGUAUGGAG	A UUUGAGA C A - AUA UUC GCC GUUCCAUGCUUC UUGCAUUC AUA GUU \\ GAG CGG CGAGGUAUGAAG AAUGUAAG UAU CGA U - UCUAAAG A G A ACU
miR-1b	UGGAAUGUAAAGAAGUAUGUAA	A GC --- AC UGGGA ACAUACUUCUUUAUAU CCAUA UGG \\ ACUCU UGUAUGAAGAAAUGUA GGUAU AUC C A A- CGA GU AL449263.5

Fig. 7 (cont.)

miR-1c	UGGAAUGUAAAGAAGUAUGUAC	
miR-1d	UGGAAUGUAAAGAAGUAUGUAUU	C GC UGAACC GCUUGGGA ACAUACUUCUUUAUAU CCAUA U CGGACUUU UGUAUGAAGAAAUGUA GGUAU G A A- CGAAUC
miR-2a-1	UAUCACAGCCAGCUUUGAUGAGC	- - A AUUUC UU GCUGGGCUC UCAAAG UGGUUGUGA AUGC CGC \\ CGAUUCGAG AGUUUC ACCGACACU UACG GCG U U G A ----- CG
miR-2a-2	UAUCACAGCCAGCUUUGAUGAGC	A C -- GAUAC AUCU AGC UCAUCAAG UGGUUGUGAUAUG \\ UAGG UCG AGUAGUUU ACCGACACUAUAC C A - CG GCAAC
miR-2b-1	UAUCACAGCCAGCUUUGAGGAGC	U UG - A C---- U CU CAAC UCUUCAAAG UGGC GUGA AUGUUG C GG GUUG AGGAGUUUC ACCG CACU UAUAAC A C CG G A AUACU A
miR-2b-2	UAUCACAGCCAGCUUUGAGGAGC	A - A UUU-- CUU UUGUGUC UUCUUCAAAG UGGUUGUGA AUG GC U AGCGCAG GAGGAGUUUC ACCGACACU UAC CG U C G A UUAUC UAU
miR-3	UCACUGGGCAAAGUGUGUCUCA	C C G U UUCA GAUC UGGGAUGCAU UUGU CAGU AUGU \\ CUAG ACUCUGUGUG AACG GUCA UACA A A A G C CUCU

Fig. 7 (cont.)

miR-4	AUAAAGCUAGACAACCAUUGA	U UU C C C GG UU UUGCAAU AGUUUC UGGU GUC AGC UUA UGAUU \ GGUGUUG UUGAAG ACCA CAG UCG AAU ACUGG U C UU A A A -- CC
miR-5	AAAGGAACGAUCGUUGUGAUAUG	UA--- C AGUUGU GC AAAGGAA GAUCGUUGUGAUAUG \ CG UUUCCUU UUAGUGACACUAUAC U CAAUA - AAUCCU
miR-6-1	UAUCACAGUGGCUGUUCUUUUU	A- C AG UAAUA UUUA UGUAGAGGGAAUAGUUGCUGUG UGUA U \ AAAU AUGUUUUUCUUGUCGGUGACAC AUAU A U CC U CU UACCA
miR-6-2	UAUCACAGUGGCUGUUCUUUUU	C UU UG C U - G UAACC AAGGGAAC C CUG UGAUAUA UA UU A GUUGG UUUUCUUG G GAC ACUAUAU AU AA A U UC GU - C C A
miR-6-3	UAUCACAGUGGCUGUUCUUUUU	A A U AAAC CAAA AGAAGGGAACGGUUGCUG UGAUGUAG UUG \ GUUU UUUUUUCUUGUCGGUGAC ACUAUAUU AAC U G - U ACUC
miR-7	UGGAAGACUAGUGAUUUUGUUGU	U U U U -- UGGUC GAGUGCAU CCGUA GGAAGAC AG GAUUU UGUUGUU \ UUUACGUG GGCAU UCUUCUG UC CUAAA ACAAUAA U C - U C UA UGGUU
miR-8	UAAUACUGUCAGGUAAAGAUGUC	CUGUUC - G C UCCUUU AAGGACAU ACAUCUU ACC GGCAG AUUAGA \ UUCCUGUG UGUAGAA UGG CUGUC UAAUCU U CCUGC- A A A CAAUAU

Fig. 7 (cont.)

miR-9	UCUUUGGUUAUCUAGCUGUAUGA	- U UAU G - GAU GCUA UGUUG CUUUGGU CUAGCU UAUGA GU A CGAU AUAAU GAAGCCA GAUCGA AUACU CA A U U UUC A G AUA
miR-10	ACCCUGUAGAUCCGAAUUUGU	CU - G U AUACU CCACGU ACC CU UAGA CCGAAUUUGUUUU A GGUGUG UGG GA AUCU GGCUUAAACAGGA G UU A G U AUUUC
miR-11	CAUCACAGUCUGAGUUCUUGC	U UCU CCC U ACU GCACUUG CAAGAACUU CUGUGA GCG GU U CGUGAGU GUUCUUGAG GACACU CGC CG A C UCU A-- - AAA
miR-12	UGAGUAUUACAUCAGGUACUGGU	UG U C - GCCUU UACGGU AGUAU ACAU AGGUACUGGU GU A GUGCCG UCAUA UGUA UUCAUGACCA CA A CA C - A ACCUA
miR-13a	UAUCACAGCCAUUUUGAUGAGU	U C - A UC-- CU UACG AACUC UCAAAG GGUUGUGA AUG GA A GUGC UUGAG AGUUUU CCGACACU UAC CU U U U A A UCAU AU
miR-13b-1	UAUCACAGCCAUUUUGACGAGU	UG- U ACU UAUU CCA UCGUUAAAAUG UUGUGA UAUG C GGU AGCAGUUUUAC GACACU AUAC A UUG C --- UAAC
miR-13b-2	UAUCACAGCCAUUUUGACGAGU	UAUU G A GCUA UU AAC CGUCAAAAUG CUGUGA UGUGGA U UUG GCAGUUUUAC GACACU AUACUU G GU-- A C ---- CA

Fig. 7 (cont.)

miR-14	UCAGUCUUUUUCUCUCUCCUA	C C C GCUU UGUGGGAG GAGA GGGGACU ACUGU \\ AUAUCCUC CUCU UUUCUGA UGAUA A U U C AAUU
miR-15a	UAGCAGCACAUAAUGGUUUGUG	GAGUAAAGUA UA GA U CCUUG GCAGCACA AUGGUUUGUG UUU \\ GGAAC CGUCGUGU UACCGACGU AAA G AUAAAAACUC UA GG A
miR-15b	UAGCAGCACAUCAUGGUUUACA	U C C A A ACA CUG AGCAGCA AU AUGGUUU CAU CU \\ GAU UCGUCGU UA UACUAAG GUA GA G C U U C - ACU
miR-16	UAGCAGCACGUAAAUAUUGGCG	AG C - A CGUUA UCUA GUCAGC UGC UUAGCAGCAC GU AAUAUUGG AGAU \\ CAGUUG AUG AGUCGUCGUG CA UUAUGACC UCUA A GA A U A ----- UUAA
miR-16	only different precursor	UC CU UA C AG AAU GU CACU AGCAGCACG AAUAUUGG GU UGA A CA GUGA UCGUCGUGU UUAUAACC CA AUU U GU UU CA A A- AUA
miR-17	ACUGCAGUGAAGGCACUUGU	GA CA- A G G - AUA GUCA AUAAUGU AAGUGCUU CA UGCAG UAG UG \\ CAGU UAUUACG UUCACGGA GU ACGUC AUC AC U GG AUG A G - U GUG
miR-18	UAAGGUGCAUCUAGUGCAGAUA	CU U C U A UGAA AG UGUU AAGG GCAU UAG GCAG UAG GU A ACGG UUCC CGUG AUC CGUC AUC CG U UC U A C - UA-- AU

Fig. 7 (cont.)

miR-19a	UGUGCAAAUCUAUGCAAAACUGA	U U -- --- AGA GCAG CC CUGUUAGUUUUGCAUAG UUGCAC UACA \\ CGUC GG GGUAGUCAAAACGUAUC AACGUG AUGU A C U UA UUG AAG
miR-19b-1	UGUGCAAAUCCAUGCAAAACUGA	UU - - UC UGUGUG CACUG CUAUGGUUAGUUUUGCA GG UUUGCA CAGC \\ GUGAU GGUGUCAGUCAAAACGU CC AAACGU GUCG A -- A U -- UCUUAU
miR-19b-2	UGUGCAAAUCCAUGCAAAACUGA	CUAC - - UUCA U ACAUUG UUACAAUUAGUUUUGCA GG UUUGCAU GCGUAUA A UGUAAU AGUGUUAGUCAAAACGU CC AAACGUG UGUAUAU U ---- A U UCGG G
miR-20	UAAAGUGCUUAUAGUGCAGGUAG	C A- G - UU GUAG ACU AAGUGCUUAUAGUGCAG UAG UG U CGUC UGA UUCACGAGUAUUACGUC AUC AU A A AA - U UG
miR-21	UAGCUUAUCAGACUGAUGUUGA	A A A U AA UGUCGGGUAGCUUAUC GACUG UGUUG CUGU G \\ ACAGUCUGUCGGGUAG CUGAC ACAAC GGUA C U - C - - UC
miR-22	AAGCUGCCAGUUGAAGAACUGU	U CC - A U CCUG GGC GAG GCAGUAGUUCUUCAG UGGCA GCUUUA GU \\ CCG CUC CGUUGUCAAGAAGUU ACCGU CGAAAU CG A U C- G - - ACCC
miR-23a	AUCACAUUGCCAGGGAUUUCC	C C - G G CUUC GG CGG UGGGG UUCCUGG GAUG GAUUUG C CC GCC ACCUU AGGGACC UUAC CUAAAC U A A U G A ACUG

Fig. 7 (cont.)

miR-23b	AUCACAUUGCCAGGGAUUACCAC	C U -- - C GUGACU GG UGC UGG GUUCCUGGCA UG UGAUUU U CC ACG ACC UAGGGACCGU AC ACUAAA G A C AU U - AUUAGA
miR-24-1	UGGCUCAGUUCAGCAGGAACAG	G G A UA UCUCAU CUCC GU CCU CUGAGCUGA UCAGU \ GAGG CA GGA GACUUGACU GGUCA U A A C C- CACAUU
miR-24-2	UGGCUCAGUUCAGCAGGAACAG	CC CG CU- AA-- UU CUCUG UCC UGC ACUGAGCUG ACACAG \ GGGAC AGG ACG UGACUCGGU UGUGUU G A- -- ACU CACA UG
miR-25	CAUUGCACUUGUCUCGGUCUGA	A AG G UU G UG ACG GGCC GUGUUG AGGC GAGAC G GCAAU CUGG C CCGG CGUGAC UCUG CUCUG C CGUUA GGUC U C AG G UU A CG CCG
miR-26a	UUCAAGUAAUCCAGGAUAGGCU	- G U U GCAG AGGCC GUG CCUCGU CAAGUAA CCAGGAUAGGCUGU G UCCGG CGC GGGGCA GUUCAUU GGUUCUAUCCGGUA U G A C - ACCC
miR-26b	UUCAAGUAAUUCAGGAUAGGUU	GA - U UC UGUG CCGG CCC AGU CAAGUAAU AGGAUAGGUUG \ GGCC GGG UCG GUUCAUUA UCUUGUCCGAC C AG C - CC CUGU
miR-27a	UUCACAGUGGCUAAGUUCCGCU	A A A U G UCCAC CUG GG GC GGGCUUAGCUGCU GUGAGCA GG \ GAC CC CG CUUGAAUCGGUGA CACUUGU CU A C C C - G GAACC

Fig. 7 (cont.)

miR-27b	UUCACAGUGGCUAAGUUCUG	AUUG UGAU U AGGUGCAGAGCUUAGCUG GUGAACAG UGG \\ UCCACGUCUUGAAUCGGU CACUUGUU GCC U GA-- UC-- U
miR-28	AAGGAGCUCACAGUCUAUUGAG	C A U ---- CC GGU CUUGCCCUC AGGAGCUCACAGUCUA UG AGUUA U UCA GGACGGGAG UCCUCGAGUGUUAGAU AC UCAGU U C G C CCUU CU
miR-29a	CUAGCACCAUCUGAAAUCGGUU	UUU C UCAAU AUGACUGAUUUC UGGUGUU AGAG \\ UAUUGGCUAAAG ACCACGA UCUU A UCU - UUAAU
miR-29b	UAGCACCAUUUGAAAUCAGUGUU	A U GU UUAAAU AGGA GCUGGUUUCA AUGGUG UUAGAU \\ UCUU UGACUAAAGU UACCAC GAUCUG A G U -- UUAGUG
miR-29c	UAGCACCAUUUGAAAUCGGuua	
miR-30a-s	UGUAAACAUCCUCGACUGGAAGC	A UC ----- A GCG CUGUAAACAUCC GACUGGAAGCU GUG A CGU GACGUUUGUAGG CUGACUUUCGG CAC G C -- GUAAA C
miR-30a-as	CUUUCAGUCGGAUGUUUGCAGC	A UC ----- A GCG CUGUAAACAUCC GACUGGAAGCU GUG A CGU GACGUUUGUAGG CUGACUUUCGG CAC G C -- GUAGA C

Fig. 7 (cont.)

miR-30b	UGUAAACAUCCUACACUCAGC	U - UCAUA AUGUAAACAUCC ACA CUCAGCUG C UGCAUUUGUAGG UGU GGGUCGGU A - A UGCGU
miR-30c	UGUAAACAUCCUACACUCUCAGC	UACU U ACA GUGGAA AGA GUAAACA CCU CUCUCAGCU A UCU CAUUUGU GGA GAGGGUCGA G UUCU C A-- AAGAAU human
miR-30d	UGUAAACAUCCCCGACUGGAAG	U U CCC GUAAGA GU GU GUAAACAUC GACUGGAAGCU C CA CG CGUUUGUAG CUGACUUUCGA A U U A-- AUCGAC chr8 human
miR-31	GGCAAGAUGCUGGCAUAGCUG	GA G C U- GAA GGAGAG GGCAA AUG UGGCAUAGC GUU C CCUUUC CCGUU UAC ACCGUAUCG CAA U UA A A UC GGG
miR-32	UAUUGCACAUUACUAAGUUGC	U - UU C GGAGAUAUUGCACAU ACUAAGUUGCAU G GU A CUUUUAUAGUGUGUG UGAUUUAACGUA C CG C - A UC G
miR-33	GUGCAUUGUAGUUGCAUUG	A UU UUCU UG CUGUGGUGCAUUGU G GCAUUGCAUG GG \ GACACUACGUGACA C UGUAACGUAC CC G C UU ---- AU
miR-99a	ACCCGUAGAUCCGAUCUUGU	A UC U G AAG CAUA ACCCGUAGA CGA CUUGUG UG U GUGU UGGGUAUCU GCU GAACGC GC G C UU C - CAG

Fig. 7 (cont.)

miR-99b	CACCCGUAGAACCGACCUUGCG	CC AC C ---- - C GGCAC ACCCGUAGA CGA CU UGCGG GG \\ CUGUG UGGGUGUCU GCU GA ACGCC CU C CC GU C ACAC G U
miR-101	UACAGUACUGUGAUAACUGA	A GUCCA UCAGUUAUCACAGUGCUG UGCU U AGUCAAUAGUGUCAUGAC AUGG U - AAAUC
miR-122a	UGGAGUGUGACAAUGGUGUUUGU	GG C UGUCC AGCUGU AGUGUGA AAUGGUGUUUG A UCGAUA UCACACU UUACCGCAAAC A AA A UAUCA woodchuck
miR-122b	UGGAGUGUGACAAUGGUGUUUGA	
miR-122a,b	UGGAGUGUGACAAUGGUGUUUG	
miR-123	CAUUAUUACUUUUGGUACGCG	A A U CGCUG C UGAC GC CAUUAUUACUU UGGUACG UGA A ACUG CG GUAAUAAUGAG GCCAUGC ACU C G C U UCAA- U
miR-124a*	UUAAGGCACGCGGUGAAUGCCA	- C A GA UAAUG CUCU G GUGUUCAC GCG CCUUGAUU U GAGA C CGUAAGUG CGC GGAAUUAA C A - G AC CAUAU

Fig. 7 (cont.)

miR-124b	UUAAGGCACGCGGGUGAAUGC	CC A GA UAAUG CUCU GUGUUCAC GCG CCUUGAUU \\ GAGA CGUAAGUG CGC GGAAUUAA U AC G AC CAUAC AC021518
miR-125a	UCCCUGAGACCCUUUAACCUGUG potential lin-4 ortholog	C C UA ---- A CUGGGU CCUGAGA CCUU ACCUGUGA GG C GGUCCG GGGUUCU GGAG UGGACACU CC G A U -- GGGA U
miR-125b	UCCCUGAGACCCUAACUUGUGA potential lin-4 ortholog	UC C A GG- U GCCUAG CCUGAGA CCU ACUUGUGA UAU U CGGAUC GGGUUCU GGA UGAACACU AUG U CA U C ACA A
miR-126	UCGUACCGUGAGUAAUAAUGC	A U CGCUG C GC CAUUAUUACUU UGGUACG UGA A CG GUAAUAAUGAG GCCAUGC ACU C C U UCAA- U
miR-127	UCGGAUCCGUCUGAGCUUGGCU	A U G G C -- AG CC GCC GCU AAGCUCAGA GG UCUGAU UC \\ GG UGG CGG UUCGAGUCU CC AGGCUA AG A C U - G U CU AA
miR-128	UCACAGUGAACCGGUCUCUUUU	UUC UAG CU U GUUGGA GGGGCCG CACUGU GAGAGGU U CGACUU CUCUGGC GUGACA CUCUUUA A UUU CAA -- C
miR-129	CUUUUUUCGGUCUGGGCUUGC	- C CU G UUCCU C GGAU CUUUUUG GGU GGGCUU CUG CU A UCUA GAAAAAC CCA CCCGAA GAC GA A U C UU G UGAU- C human

Fig. 7 (cont.)

miR-130	CAGUGCAAUGUUAAAAGGGC	- C A GUCUAAC GA GCUCUUUU ACAUUGUGCU CU \\ CU CGGGAAAA UGUAACGUGA GA G A U C GCCAUGU
miR-131	UAAAGCUAGAUAACCGAAAGU	G C G U A GUU UUAU UUUGGUUAUCUAGCU UAUGAG GU U CAA AAUG AAGCCAAUAGAUCGA AUACUU UG U A A A C G
miR-132	UAACAGUCUACAGCCAUGGUCGU	A UUC G- G GGGC ACCGUGGCU GAUUGUUACU UGG \\ CCCG UGGUACCGA CUGACAAUGG GCC A C CAU AG A
miR-133	UUGGUCCCCUUCAACCAGCUGU	A AA U A GCCUC GCUA AGCUGGU AA GG ACCAAAUC U CGAU UCGACCA UU CC UGGUUUAG U G AC C C GUAAC
miR-134	UGUGACUGGUUGACCAGAGGGA	GU U A- G GCGU AC AGGGU GUGACUGG UG CCA AGGG GC \\ UCCCA CACUGAUC AC GGU UCCC UG U AC C CG G ACU- UC
miR-135	UAUGGCUUUUUAUUCCUAUGUGAA	UU UUCUAU CUAUGGCUUU AUUCCUAUGUGA \\ GGUGCCGAGG UAGGGAUAUACU U U- CGCUCG
miR-136	ACUCCAUUUGUUUUGAUGAUGGA	C UUU UUCU GAGGACUC AUUUG UGAUGAUGGA \\ CUUCUGAG UAAAC GCUACUACCU U - UCU CGAA

Fig. 7 (cont.)

<table>
<tr><td>miR-137</td><td>UAUUGCUUAAGAAUACGCGUAG</td><td><pre>
        G   G   A   - GA
CUUCGGU ACG GUAUUCUUGGGUGG UAAUA CG   \
GGAGCUG UGC CAUAAGAAUUCGUU AUUGU GC   U
        A   G   -   U AU
</pre></td></tr>
<tr><td>miR-138</td><td>AGCUGGUGUUGUGAAUC</td><td><pre>
     --   UCA   AC-   C CG
CAGCU GGUGUUGUGAA   GGCCG   GAG AG C
GUUGG CCACAGCACUU   UCGGC   UUC UC A
     GA   UA-   CCA   - CU
</pre></td></tr>
<tr><td>miR-139</td><td>UCUACAGUGCACGUGUCU</td><td><pre>
G   -   U A   GUGGC
GU UAUUCUA CAG GC CGUGUCUCCAGU   \
CA AUGAGGU GUC CG GCGCAGAGGUCG   U
-          U   C -   GAGGC   human
</pre></td></tr>
<tr><td>miR-140</td><td>AGUGGUUUUACCCUAUGGUAG</td><td><pre>
    - A   A   UU UC
CCUG CC GUGGUUUUACCCU UGGUAGG   ACG A
GGAC GG CACCAAGAUGGGA ACCAUCU   UGU U
    A -   C   -- CG
</pre></td></tr>
<tr><td>miR-141</td><td>AACACUGUCUGGUAAAGAUGG</td><td><pre>
   U   --   U   AU   GAAG
GGG CCAUCUU   CCAG GCAGUGUUGG   GGUU   \
CCC GGUAGAA   GGUC UGUCACAAUC   UCGA   U
   -         AU   -   C-   AGUA
</pre></td></tr>
<tr><td>miR-142s</td><td>CAUAAAGUAGAAAGCACUAC</td><td><pre>
AC-   A   UAA--- G
   CCAUAAAGUAG AAGCACUAC   CA C
   GGUAUUUCAUC UUUGUGAUG   GU A
GUA   C   UGGGAG C
</pre></td></tr>
<tr><td>miR-142as*</td><td>UGUAGUGUUUCCUACUUUAUGG</td><td><pre>
AC-   A   UAA--- G
   CCAUAAAGUAG AAGCACUAC   CA C
   GGUAUUUCAUC UUUGUGAUG   GU A
GUA   C   UGGGAG C
</pre></td></tr>
</table>

Fig. 7 (cont.)

new	AUAAGACGAGCAAAAAGCUUGU	G G C GG C AU UGAC GGCGAGCUUUU GC CG UUAUAC UG \\ ACUG UUGUUCGAAAA CG GC AAUAUG AC G G A A AG C UC AL049829.4
miR-143	UGAGAUGAAGCACUGUAGCuca UUAGAUGAAGCACUGUAG	G G U - AG CCUGAG UGCAGUGCU CAUCUC GG UC U GGACUC AUGUCACGA GUAGAG CU AG U G A U G GG AC008681.7
miR-144	UACAGUAUAGAUGAUGUACUAG	G A A- GU GGCUGG AUAUCAUC UAUACUGUA GUUU G CUGAUC UGUAGUAG AUAUGACAU CAGA A A - CA GU
miR-145	GUCCAGUUUUCCCAGGAAUCCCUU	C UC U C UGGAUG CUCA GG CAGU UU CCAGGAAUCCCU \\ GAGU UC GUCA AA GGUCCUUAGGGG C - UU U A UAGAAU
miR-146	UGAGAACUGAAUUCCAUGGGUUU	CU C AUAUC AGCU GAGAACUGAAUU CAUGGGUU A UCGA UUCUUGACUUAA GUGUCCAG A C- A ACUGU
miR-147	GUGUGUGGAAAUGCUUCUGCC	A- CAA ACA--- GA AAUCUA AGA CAUUUCUGCACAC CCA \\ UUAGAU UCU GUAAAGGUGUGUG GGU C CG UC- ACCGAA AU human
miR-148	UCAGUGCACUACAGAACUUUGU	- A- CC - AGU GAGGCAAAGUUCUG AG CACU GACU CUG \\ CUCUGUUUCAAGAC UC GUGA CUGA GAU A A AC -- A AGU human

Fig. 7 (cont.)

miR-149	UCUGGCUCCGUGUCUUCACUCC	G C G A GUG G GGCUCUG CUC GU UCUUC CUCCC UUU U UCGGGGC GAG CA GGAGG GAGGG GAG C G A G - AG- C
miR-150	UCUCCCAACCCUUGUACCAGUGU	AC U UG- UG CCCUGUCUCCCA CCU GUACCAG CUG \\ GGGAUAGGGGGU GGA CAUGGUC GAC C CC - CCA UC
miR-151	CUAGACUGAGGCUCCUUGAGGU	C CA UGUCU CCUG CCUCGAGGAGCU CAGUCUAGUA \\ GGAC GGAGUUCCUCGG GUCAGAUCAU C A A- CCCUC
miR-152	UCAGUGCAUGACAGAACUUGG	G A CC CGG C CCGGGCCUAGGUUCUGU AU CACU GACU GCU U GGCCCGGGUUCAAGACA UA GUGA CUGA CGA G G C -- --- G
miR-153	UUGCAUAGUCACAAAAGUGA	- GU A- AAU CAGUG UCAUUUUUGUGAU UGCAGCU GU \\ GUUAC AGUGAAAACACUG ACGUUGA CG A U AU CC AGU
miR-154	UAGGUUAUCCGUGUUGCCUUCG	U - CCU-- UUU GAAGAUAGGUUA CCGUGU UG UCGC \\ UUUUUAUCCAGU GGCACA AC AGUG A U U UAAGC UUU
miR-155 [BIC-RNA]	UUAAUGCUAAUUGUGAUAGGGG	U U A UUGGCC CUGUUAAUGCUAAU G G UAGGGGUU \\ GACAAUUACGAUUG U C AUCCUCAG U - C - UCAGUC

Fig. 7 (cont.)

name	sequence	structure
miR-C1	AACAUUCAACGCUGUCGGUGAGU	U A U CU A GGGAUUCA CCA GG ACA UCAACG GUCGGUG GUUU \\ GGU CC UGU AGUUGC CAGCCAC CAAA A U A C -- - AAAACAAA
miR-C2	UUUGGCAAUGGUAGAACUCACA	UU UGG UCA UAAGGU ACCAU UUGGCAA UAGAAC CACCGG A UGGUA AACCGUU AUCUUG GUGGCC A UC CAG --- CAGGGU
miR-C3	UAUGGCACUGGUAGAAUUCACUG	G AC-- GA -- AC CUGU UAUGGC UGGUA AUUCACUG UGA A GACA AUACCG GCCAU UAAGUGAC ACU G A GGAA -- UG CU
miR-C4	CUUUUUGCGGUCUGGGCUUGUU	- C CU G UUUU C UGGAU CUUUUUG GGU GGGCUU CUG CU G AUCUA GAAAAAC CCA CCCGAA GAC GA A U C UU G UGAU C
miR-C5	UGGACGGAGAACUGAUAAGGGU	U C AG - UG CCU UCCUUAUCA UUUUCC CCAGC UUUG A GGA GGGAAUAGU AAGAGG GGUUG GAAU C U C CA U CU
miR-C6	UGGAGAGAAAGGCAGUUC	A G AU UC AGGGAUUGGAG GAAAG CAGUUCCUG GG C UUCCUGGUCUC CUUUC GUCGGGGAC CC C - G -- UC

Fig 7 (cont.)

name	sequence	structure
miR-C7	CAAAGAAUUCUCCUUUUGGGCUU	U UU UCUCAU ACUUUCCAAAGAAUUC CCUU GGGCUU U UGAAGGGUUUUUUAAG GGAA CCCGAA U U U- UUUUAU
miR-C8	UCGUGUCUUGUGUUGCAGCCGG	A A C CGCUGC UC GGCU CAACACAGGAC CGGG U GG CCGA GUUGUGUUCUG GCUC C - C U CCCAGU
miR-C9	UAACACUGUCUGGUAACGAUGU	- C UU UUG GGGCAUC UUACCGGACAGUG UGGA UC \\ CUUGUAG AAUGGUCUGUCAC AUCU AG G C A C- UUC
miR-C10	CAUCCCUUGCAUGGUGGAGGGU	CA UC GU UGAGCUC UCU CA CCUUGCAUG GGAGGG U AGG GU GGGACGUAC CCUCCC C AC UU AC CAAAAGU
miR-C11	GUGCCUACUGAGCUGACAUCAGU	G G A UA UCUCAU CUCC GU CCU CUGAGCUGA UCAGU \\ GAGG CA GGA GACUUGACU GGUCA U A A C C- CACACU
miR-C12	UGAUAUGUUUGAUAUAUUAGGU	U- UA--- UU CUGUG GAUAUGUUUGAUAUAU GGUUG \\ GACAU UUAUACGAACUAUAUA CUAAU A CC UCAAC UU

Fig. 7 (cont.)

name	sequence	structure
miR-C13	CAACGGAAUCCCAAAAGCAGCU	C C AA UU - C AGCGGG AACGGAAUCC AA GCAGCUG GU CU C UCGUCC UUGCUUUAGG UU CGUCGAC UA GA A C - CA CU C G
miR-C14	CUGACCUAUGAAUUGACA	C - A UGCUCUC UGACCUAUG AAUUG CAGCCAG G ACUGGAUAC UUAAC GUCGGUC U - C C UCCCCUC
miR-C15	UACCACAGGGUAGAACCACGGA	- G A UU UC UCCUG CCG UGGUUUUACCCU UGGUAGG ACG A AGGAC GGC ACCAAGAUGGGA ACCAUCU UGU U A - C -- CG
miR-C16	AACUGGCCUACAAAGUCCCAG	A U C A A AGU GAG GCUGGG CUUUG GGGC AG UGAG G CUC UGACCC GAAAC UCCG UC ACUU U C U A G A GAC
miR-C17	UGUAACAGCAACUCCAUGUGGA	U A G -- U AUCGGG GUAACAGCA CUCCAU UGGA CUG G UAGUCU CAUUGUCGU GAGGUG ACCU GGC C U C - UA U
miR-C18	UAGCAGCACAGAAAUAUUGGC	U A- UG GAA AGCAGCACAG AAUAUUGGCA GG G UCGUCGUGUC UUAUAACCGU CU U GG -- GAG

Fig 7 (cont.)

name	sequence	structure
miR-C19	UAGGUAGUUUCAUGUUGUUGG	A A C GGCCUGGG GUGAAUU GGU GUUU AUGUUGUUG U CACUUAG CCA CAAA UACAACAAC U C C U ACAAGUCU
miR-C20	UUCACCACCUUCUCCACCCAGC	C A CA GA - A GGCUGUGC GGGU GAGAGGG GUGG GGU AAG G CCGGUACG CCCA CUCUUCC CACU CCA UUC C A C AC UC C U
miR-C21	GGUCCAGAGGGGAGAUAGG	G - C G U UUCCUG UCAUU G UC A AGGGGAGA AGG U AGUAA U AG U UCUCUUCU UCC G A A A A - UUUUUA
miR-C22	CCCAGUGUUCAGACUACCUGUU	AAC U C U G--- G GCC CCAGUGU CAGACUAC UGU CA GAG \\ CGG GGUUACA GUCUGAUG ACA GU CUC C AUU C - U GUAA U
miR-C23	UAAUACUGCCUGGUAAUGAUGAC	GGC - C UAGUG GCCGU CAUC UUACUGGGCAG AUUGGA U CGGCA GUAG AAUGGUCCGUC UAAUCU C --- U A CUAGU
miR-C24	UACUCAGUAAGGCAUUGUUCU	U U UUC A UACCUUAC CAG AAGGCAUUGUUC UAU U AUGGGAUG GUC UUCCGUGACAAG AUA U U U UAA A

Fig.7 (cont.)

name	sequence	structure
miR-C25	AGAGGUAUAGCGCAUGGGAAGA	U A- UG C GUUCC UUUUCCUAUGC UAUACUUCUU UGGAU \\ CGAGG AGAAGGGUACG AUAUGGAGAA AUCUG U U CG -- G
miR-C26	UGAAAUGUUUAGGACCACUAG	C U G A C U GGUC AGUGGUUCU GACA UUCA CAGUU UG \\ CCAG UCACCAGGA UUGU AAGU GUUAA AC A A U A - C G
miR-C27	UUCCCUUUGUCAUCCUAUGCCUG	U A U GAGAAUA UGGAC UCCCUUUGUC UCCUA GCCU \\ ACUUG AGGGAAACGG AGGGU CGGA U C A - GGAAGUA
miR-C28	UCCUUCAUUCCACCGGAGUCUG	UC C UCUUA CUCUUG CUUCAUUCCAC GGAGUCUG U GAGGAC GAAGUGAGGUG CUUUAGAC G UC A CAACC
miR-C29	GUGAAAUGUUUAGGACCACUAGA	U C U G A C U GCC GGUC AGUGGUUCU GACA UUCA CAGUU UG \\ CGG CCAG UCACCAGGA UUGU AAGU GUUAA AC A C A U A - C G
miR-C30	UGGAAUGUAAGGAAGUGUGUGG	- C U AUAUC CCAGG CCACAUGCUUCUUUAUAU C CAUAG \\ GGUUU GGUGUGUGAAGGAAUGUA G GUAUC U U A - ACGAC

Fig 7 (cont.)

name	sequence	structure
miR-C31	UACAGUAGUCUGCACAUUGGUU	AUC U C ------- G GCC CCAGUGU CAGACUAC UGU UCAG A CGG GGUUACA GUCUGAUG ACA GGUC G AUU C - UGUACAG G
miR-C32	CCCUGUAGAACCGAAUUUGUGU a miR-10 variant	A G C UG- AC UAUAU CCCU UAGAA CGAAUUUGUG GU C AUAUA GGGG AUCUU GCUUAGACAC UA C A - A UGA CA
miR-C33	AACCCGUAGAUCCGAACUUGUGA A a miR-99a variant	A C C A C AU CACA ACC GUAGAU CGA CUUGUG UG U GUGU UGG UAUCUG GUU GAACAC AC C A A U C - GU
miR-C34	GCUUCUCCUGGCUCUCCUCCCUC	C U UUG - GGAG AAGG AGGGG GAGGGG CGGGAGGAGC CGGGC G UUCC UCUCC CUCCUC GUCCUCUUCG GUUCG C - - UCG C GCGU

Fig. 7 (cont)

name	human	C.elegans	mouse								Drosophila	fugu fish	zebrafish
			liver	small intes	colon	cerebellum	cortex	midbrain	heart	spleen			
let-7a-1	AC007924 chr9 AC087784 chr 17 identical precursor		num.hits in trace data, 3 families of similar precursors		found		nearly identical precursor	found					
let-7a-2	AP001359 chr11						nearly identical precursor						
let-7a-3	AL049853 chr22	AF274345 chrX with diff. precursor									AE003659 diff. Precursor		
let-7b	AL049853 chr22		nearly identical precursor			nearly ident precursor trace#48311003		found		EST AI481799.1 spleen = cerebellum (mammary)		with slightly diff precursor	
let-7c	AP001667 chr21		identical and diff. precursors			num.genomic hits, ident precursor;diff precursor -> EST AI614897	numerous genomic hits	found					
let-7d	AC007924.3 chr9 AC087784 chr17 identical				found	trace#83587042 nearly ident prec	trace#83587042 nearly ident prec	found	found				
let-7e	AC018755 chr19							found		FOUND			
let-7f-1	AC007924 chr9 AC007704 chr17					ident precursor genomic DNA		found		found			
let-7f-2	AL592046 chrX					ident. precursor in mmtrace 18713911							
let-7g	precursor ident. to mouse in AC092045.2 chr3					genomic hits,no EST		found					
let-7h							found in cortex,no db hit						

Fig. 7 (cont.)

let-7i	precursor ident. to mouse [AL117383.19]; also AC048341.22					found, supported by EST BB661268	found						
miR-1											2L, AE003667		
miR-1b	AL449263.5 chr20 nt1-21	U97405.1 nt 1-21 (22G)	no mouse hit (only nt1-21)					found	found				
miR-1c									found, but no db hit				
miR-1d	AL449263.5 chr20 nt1-22 (23G)							found	trace hits (nt1-23) trace#91 523974				BF157601.1 with C23 (diff. precursor)
miR-2a-1											2L, AE003663		
miR-2a-2											2L, AE003663		
miR-2b-1											2L, AE003620		
miR-2b-2											2L, AE003663		
miR-3											2R, AE003795		
miR-4											2R, AE003795		

Fig. 7 (cont.)

miR-5											2R, AE003795		
miR-6-1											2R, AE003795		
miR-6-2											2R, AE00379		
miR-6-3											2R, AE00379		
miR-7	AC003791 chr19 diff.precursor; EST BF373391 again different		not cloned, but mouse EST predicts precursor similar to human								2R, AE003791		
miR-8											2R, AE003805		
miR-9	AC005316 chr15 AC026701 chr5 each with diff. precursor					AF155142.1 chr19 diff prec,sligh.diff prec.s in trace hits		found			3L, AE003516	2diff precurs scaffold_ 3868 and 2417	
miR-10	AF287967 chr11 (HOX B4/B5)		not found, but AC011194 chr.11 predicts diff. precursor								AE001574		
miR-11											3R, AE003735		
miR-12											X, AE003499		
miR-13a											3R, AE003708		

Fig. 7 (cont.)

miR-13b-1											3R, AE003708		
miR-13b-2											X, AE003446		
miR-14											2R, AE003833		
miR-15a	13, AC069475							found	trace#72 137197 prec slig diff				
miR-15b									trace#79 105069				
miR-16	13, AC069475 interesting leukemia locus					genomic hits with 2 slightly diff precur.trace#502 93836,78368680		found					AL606727 diff precurs
miR-16	3, NT_005740.6			several trace,near ly ident precursor	found		found trace#7910506 9;nearly ident prec. as in human		found				
miR-17	13, AL138714												
miR-18	13, AL138714												
miR-19a	13, AL138714												
miR-19b-1	13, AL138714									found			G46757 with a U9C

Fig. 7 (cont.)

miR-19b-2	X, AC002407												
miR-20	13, AL138714				found								
miR-21	17, AC004686			AL604063 chr11, nearly ident precursor	found				found	found			
miR-22	several highly similar ESTs: AW961681 shown		cDNAs from var. tissues, identical precursor	AK008813 cDNAs, same precursor		AK008813 (cDNA), prec ident to human		found	found				
miR-23a	19, AC020916								found trace#62 540691 prec sli diff			three hits in db	
miR-23b	XM_072557.1 chr9, also human ESTs, prec nearly ident to mouse						EST AW124037 hypothal, EST AI848465 cerebellum						
miR-24-1	9, AF043896				found		found. EST AI286629 (thymus): nearly ident. to miR-24-1; EST AA111466 (whole embryo) different precursor	found	found				
miR-24-2	19, AC020916												
miR-25	7, AC073842 second ident. copy found in chr7		predicted in mouse (EST AI595464), but not cloned										G46757 similar precursor
miR-26a	3, AP000497						AC055818.9, trace#88471973 precursor diff. from human	found				Scaffold_4097 different precursor	
miR-26b	2, AC021016		found			found, trace#69866494, slight. diff precursor							

Fig. 7 (cont.)

miR-27a	19, AC020916				found	found, but no db hit	found,but no db hit for mouse	found	found	found			
miR-27b	XM_098943.1 chr9 identical precursor							found,maps to chr 13 MGSC mmtrace 44671617					
miR-28	3, AC063932												
miR-29a	7, AF017104 second ident.copy found in chr7 CLUSTER,this cluster also consvd in mouse: AC024913.32				found, AC024913.32	found, mmtrace#23467334	nearly ident precursor trace#23467334,EST AC024913.32		trace, EST, nearly ident prec				
miR-29b	AL035209.1 chr1 CLUSTER of miR-29-b and 29-c; miRNA similar to miR-83				found		AC024913.32;diff precursor in EST BG342396 (retina)	found	FOUND			Scaffold_17670.(A third copy)	
miR-29c							found	found	found, supportd by ESTs			Scaffold_17670 has two copies of this RNA	
miR-30a-s	nearly ident fold in AL035467.23 chr6			found;ESTs ,trace68023889 all with 22G			found	found		found			
miR-30a-as	6, AL035467					found with diff. precursor in trace #85261735							
miR-30b	human AF159227.6 chr8,different precursor					trace#72329251	found			found		Scaffold_3483,diff precursor	
miR-30c	AL136164.8 chr.6 supported by ESTs (BF594736.1)					found,but no db hit for mouse			found	found			

Fig. 7 (cont.)

miR-30d	AF159227.5 chr8								found, but no mouse db hit			Scaffold 3483,diff fold	
miR-31	9, AL353732												
miR-32	9, AL354797												
miR-33	22, Z99716												G44780 with diff.precursor
miR-99a	AP000962.2 chr21,ident to mouse;[similar to miR-10 and miR-51]						trace#48910714						
miR-99b	AC018755.3 chr.19; [similar to miR-10 and miR-51]					mmtrace #92340982							
miR-101	AL158147.17 chr9 diff precursor					AK021368.1 cDNA eyeball	found					U53213.1 T.fluviatilis	
miR-122a			abundant but no db hit,except woodchuck X13234										
miR-122b													
miR-122a,b													
miR-123			genomic hits (trace#6108147), no EST									Scaffold_3295	

Fig. 7 (cont.)

miR-124a*	nearly ident. precursor in chr8(AC021518) chr20(AL096828)	found in Z72504.1 chrIV intron,diff precursor		found		most abundant in cereb.,genomic hits (trace#21097008, 11737241).	most abundant;several trace hits;precurs=cerebellum	found			slightly diff precursor AC009251 chr2L		
miR-124b	AC021518 chr8,nearly ident chr20 AL096828.29					found, but no db hit	found						
miR-125a	ident precur in AC018755.3 chr 19					genomic hits trace#33921945, 48262259 and more	found						
miR-125b	AP001359.4 chr11 AP001667.1 chr21(chr21 like mouse)						trace#83985705	found with A22U			found in AC006590.11 with diff fold	Scaffold_2358	
miR-126						mmtrace#3521597 and more			found			with diff precursScaffold_3295	
miR-127	human AL117190.6 chr.14 same precurs as in mouse					hit in trace#79514537							
miR-128	ident in AC016742.10 chr 2;diff prec in AC016943.7 chr.3					genomic hit trce#51670230	found	found				Scaffold_828,diff prec	
miR-129	human AC018662.3 chr7					found, but no db hit							
miR-130						mmtrace 68479278							with diff fold AC091299.2
miR-131	AC005317.2 chr 15 sligh.diff precursor,but AC026701.6 chr 5 ident					several trace hits,mouse AF155142	found						
miR-132	AL137038.5 chr17 prec sligh.diff from mouse					trace hit#86984641							

Fig. 7 (cont.)

miR-133	AL391221.15 chr6 diff. Precursor(ident to rat L33722.1)						found, trace#62407955		found		AC093440.1 diff. Precursor	Scaffold_1049;precu nearly like mouse	
miR-134	AL132709.5 chr14 similar precursor						trace#64620311						
miR-135	AC092045.2 chr3 AC018659.35 chr12 (ident or simil to mouse)						trace#71495235,ESTBF780995.1(kidn.,spleen)(=chr3human)			found		Scaffold_2125 with similar precurs	
miR-136	AL117190.6 chr14 ident to mouse						trace#86071753						
miR-137	AC027691.1 chr1 ,ident to mouse,nearly ident fish						trace#89774543,EST (hypothal)AI852436.1,ident.					Scaffold_18244 nearly ident to mouse/man	
miR-138	AC006058.1 chr3 precursor diff						mouse EST BB528620.2						
miR-139	AP003065.2 chr11						found, but no mouse hit						
miR-140	AC026468.8 chr.16,precursor nearly ident,			several trace hits; trace#10530393									
miR-141	AC006512.12 chr12,precursor slightli diff			AC002397 chr6						found			
miR-142s	AC004687.1 chr17 BCL3/myc translocation locus,like mouse			found	found								
miR-142as*				several EST AI153235	found					found			

Fig. 7 (cont.)

new	AL049829.4 chr14								found but no db hit				
miR-143	AC008681.7 chr5						found, but no db hit	found	found	found			
miR-144	XM_064366.1 precursor nearly ident				found				EST AA290206 .1,trace 2143909				
miR-145	AC008681.7 chr5 GG->GA;precur nearly like mouse, see 2 positions above								found EST BF163348 .1 lung			Scaffold 934 similar	
miR-146	AC008388.7 chr5 diff precursor								trace#34 639321				
miR-147	AL592549.7									found			
miR-148	AC010719.4									found, no db hit			
miR-149									trace#85 955550				
miR-150					trace#8472 1065,10352 801								
miR-151					trace#8845 6669								
miR-152	human chr 17 AC004477.1, nearly identical				found in colon,supportd.by trace#83700445;close match MGSC in chr18 (additional 14C unlikely, not supported by trace and								

Fig. 7 (cont.)

miR-153	AC006372.2 chr7 ident.precursor							found sever. mmtrace 87010874					
miR-154	AL132709.5 chr14 nearly identical precursor							found sever. mmtrace 86715639					
miR-155 (BIC-RNA)	human BIC RNA:AF402776.1 (has U12C)				found;chr 16 mouse								

Fig. 7 (cont.)

name	human	mouse							Drosophila	fugu fish	zebrafish
		spleen	eye	kidney	testes	lung	thymus	skin			
miR-C1	with different precursors in chr9 AL158075.11,chr1 AL136321.5		mouse trace #76647842			found		found		scaffold_ 1819	
miR-C2	chr7 AC084864.2 similar precursor		mouse trace #88841093							scaffold_ 967	AL590150.2
miR-C3	chr7 AC084864.2 ident.precursor		trace #86029980							scaffold_ 967	AL590150.2
miR-C4	similar precurs.in chr7 AC018662.3		trace #13885686		found						
miR-C5	chr15 AC069082.9		trace #87318220						found	scaffold_ 3671	
miR-C6	chr22 AC005664.2 ident.precursor		chr16 AC012526.32								
miR-C7	chr1 AL512443.7 similar prec.		trace #86694995								
miR-C8				found, trace #51673384							
miR-C9				found, trace #78964803						scaffold_ 2210, diff. precursor	
miR-C10	chrX AF222686.1 nearly ident. precursor			found, trace #61928192							
miR-C11	chr9 XM_098943.1 has C17U;prec.nearly identical to mouse			found,cDNA AI286629.1, has C17U							
miR-C12				found,trace#71 760450						scaffold_ 2294	
miR-C13		found		found,trace #88722637							

Fig. 7 (cont.)

name	human	mouse							Drosophila	fugu fish	zebrafish
		spleen	eye	kidney	testes	lung	thymus	skin			
miR-C14	chr11 AC000159.6			found, but no db hit							
miR-C15	chr16 AC026468.6 nearly ident.precursor			EST BI687377.1, several trace						scaffold_ 2083	
miR-C16	chr17 AC003101.1, similar precursor			found,trace#95 55103						scaffold_ 246	
miR-C17	chr11 AC000159.6, chr1 AC103590.2; diff.prec.			found, trace #87796602						scaffold_ 152	
miR-C18				found, trace #47823768 (close to miR-16)		found		found			
miR-C19	chr17 AC009789.21 cloned from human cell line only			similar precursor in mouse chr11 AC011194.15						scaffold_ 18334	
miR-C20	chr1 AL355310.19 cloned from human cell line only										
miR-C21	chr3 AC063952.15 cloned from human cell line only										
miR-C22	chr19 AC007229.1; chr1 AL137157.7 similar precursor; cloned from human cell line only									scaffold_ 8399	
miR-C23					trace #72257777	found				scaffold_ 2210	
miR-C24					trace #69879879						
miR-C25					trace #49754566						
miR-C26	AL136001 ident. precursor				trace #11977216						

Fig. 7 (cont.)

name	human	mouse							Drosophila	fugu fish	zebrafish
		spleen	eye	kidney	testes	lung	thymus	skin			
miR-C27	chr9 AL159990.12 identical precursor		trace #91503159							scaffold_ 725	
miR-C28	XM_036612.4, precursor very similar							XM_149012.1		scaffold_ 13664	
miR-C29	chr14 AL136001.6 nearly identical precursor							trace #18453604			
miR-C30	chr6 AL391221.15 similar precursor							trace #84055510			
miR-C31	chr9 AC006312.8							trace #89079710		scaffold_ 5830	
miR-C32								U77364.1, intronic location Hoxd4 gene		scaffold_ 82	
miR-C33								trace #84780544		scaffold_ 15612	
miR-C34							trace# 72109322				

SEQUENCE LISTING

<110> Max-Planck-Gesellschaft zur Förderung der Wissensc

<120> Identification of novel genes encoding for small
temporal RNA

<130> 26250PWO_DR

<140>
<141>

<150> EP 01 123 453.1
<151> 2001-09-28

<150> EP 02 006 712.0
<151> 2002-03-22

<150> EP O2 016 772.2
<151> 2002-07-26

<160> 217

<170> PatentIn Ver. 2.1

<210> 1
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 1
tactatacaa cctactacct caatttgcc 29

<210> 2
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 2

actatgcaac ctactacctc t 21

<210> 3
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 3
actatacaac ctcctacctc a 21

<210> 4
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 4
tggtgtttcc gcccgggaa 19

<210> 5
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 5
tggaatgtaa agaagtatgg ag 22

<210> 6
<211> 23
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
Oligonucleotide

<400> 6
gctcctcaaa gctggctgtg ata 23

<210> 7
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 7
tgagacacac tttgcccagt ga 22

<210> 8
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 8
tcaatggttg tctagcttta t 21

<210> 9
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 9
catatcacaa cgatcgttcc ttt 23

<210> 10
<211> 22

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 10
aaaaagaaca gccactgtga ta 22

<210> 11
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 11
tggaagacta gtgattttgt tgt 23

<210> 12
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 12
gacatcttta cctgacagta tta 23

<210> 13
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 13
tcatacagct agataaccaa aga 23

<210> 14
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 14
acaaattcgg atctacaggg t 21

<210> 15
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 15
gcaagaactc agactgtgat g 21

<210> 16
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 16
accagtacct gatgtaatac tca 23

<210> 17
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:

Oligonucleotide

<400> 17
actcgtcaaa atggctgtga ta 22

<210> 18
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 18
taggagagag aaaaagactg a 21

<210> 19
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 19
tagcagcaca taatggtttg t 21

<210> 20
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 20
gccaatattt acgtgctgct a 21

<210> 21
<211> 22
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 21
tacaagtgcc ttcactgcag ta 22

<210> 22
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 22
tatctgcact agatgcacct ta 22

<210> 23
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 23
tcagttttgc atagatttgc aca 23

<210> 24
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 24
tacctgcact ataagcactt ta 22

<210> 25
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 25
tcaacatcag tctgataagc ta 22

<210> 26
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 26
acagttcttc aactggcagc tt 22

<210> 27
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 27
ggaaatccct ggcaatgtga t 21

<210> 28
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 28
ctgttcctgc tgaactgagc ca 22

<210> 29
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 29
tcagaccgag acaagtgcaa tg 22

<210> 30
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 30
agcctatcct ggattacttg aa 22

<210> 31
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 31
agcggaactt agccactgtg aa 22

<210> 32
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 32
ctcaatagac tgtgagctcc tt 22

<210> 33
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 33
aaccgatttc agatggtgct ag 22

<210> 34
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 34
gctgcaaaca tccgactgaa ag 22

<210> 35
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 35
cagctatgcc agcatcttgc ct 22

<210> 36
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 36
gcaacttagt aatgtgcaat a 21

<210> 37
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 37
tgcaatgcaa ctacaatgca cc 22

<210> 38
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 38
ctccatactt ctttacattc ca 22

<210> 39
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 39
gctgagtgta ggatgtttac a 21

<210> 40
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 40
gcttccagtc gaggatgttt aca 23

<210> 41
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 41
cgcaaggtcg gttctacggg tg 22

<210> 42
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide

<400> 42
tcagttatca cagtactgta 20

<210> 43
<211> 23
<212> DNA
<213> Artificial Sequence